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, 2,3], Mitsuaki Matsui[4], Azusa Iwamoto[5], Bunsreng Taking[7], Sharon E Cox[4,8,9], Lindsay J Hall[1,2,3,10].

[1]Food, Microbiome & 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

The gut microbiota of infants in low-middle income countries like Cambodia remain underrepresented in microbiome research. This study aimed to explore the faecal gut microbiota composition and faecal cytokine profiles in a cohort of infants living in a rural province of Cambodia and explore the impact of sample storage conditions and infant environment on microbiota composition. Faecal samples collected at three time points from 32 infants (96 samples in total) after 7 months of age were analysed using 16S rRNA amplicon sequencing to determine the composition of the microbiota. Bacterial strains were isolated and subjected to whole genome sequencing and genomic analysis and concentrations of faecal cytokines were also measured. Initially, we compared the effects of two sample collection methods due to the challenges of faecal sample collection and storage in a rural location. Storage of faecal samples in a DNA preservation solution retained a greater abundance of *Bacteroides*. Analyses of microbiota composition of samples stored in DNA preservation solution indicated that *Bifidobacterium* was the most abundant genus with *Bifidobacterium longum* the most abundant species, particularly in breastfed infants. Most infants had detectable pathogenic taxa indicating frequent pathogen exposure, with *Shigella* and *Klebsiella* more abundant in infants with recent diarrheal illness. We did not detect antibiotic-associated perturbations in the gut microbiota, and no associations were found between the gut microbiota and infant growth. Genomic analysis of isolated strains indicated the presence of gene clusters encoding the ability to digest human milk oligosaccharides in *B. longum* and *Bifidobacterium breve* isolates. The presence of antibiotic-resistant genes was also identified in potentially pathogenic species, as well as in beneficial genera including *Bifidobacterium*. Faecal cytokine analysis showed higher concentrations of Interlukin-1alpha and vascular endothelial growth factor in breastfed infants, which may influence the infant gut mucosal immune system. This study provides insights into an underrepresented population of rural Cambodian infants, emphasising the impact of pathogen exposure and breastfeeding on gut microbiota composition and faecal immune profiles.

Introduction

Research on the gut microbiota has primarily focussed on populations in high income countries (e.g. Europe and North America) (Abdill et al., 2022), and the problems and limitations this bias is introducing to microbiome research is gaining increasing attention (Browne et al., 2024). Although more recent studies have increased representation of populations in low- and low-middle income countries (LMICs), infant studies lag behind those working on adult cohorts (Porras and Brito 2019). This is important for defining what is a 'healthy' human gut microbiota (Lackey et al., 2019) across individuals in LMICs (Porras and Brito, 2019), and because optimal infant development is particularly important for infants from LMICs due to increased risk of (mal)nutrition and infection. While a small proportion of microbiota samples studied globally so far have come from South Asia, including India and Bangladesh (Abdill et al., 2022), research into the gut microbiota of people living outside these countries remains very limited. Moreover, due to the logistics of sample collection and storage in rural areas, research in LMICs has often focused on individuals in urban areas, neglecting key infant populations. Cambodia in particular has a scarcity of microbiota studies and given that this country has such a high prevalence of infants classified as stunted/wasted, and has amongst the highest infant mortality rates in Southeast Asia, with infections accounting for many of these deaths (Cambodia Demographic and Health Survey, 2022), new studies targeting Cambodian infant populations are needed. Previous research focusing on the gut-associated microbiota of people in Cambodia has been limited with studies focusing on particular bacterial pathogens, such as antimicrobial resistance in *Enterobacterales* in faecal samples from babies and children (Auguet et al., 2021), the epidemiology of antibiotic resistance in *Escherichia coli* and *Klebsiella pneumoniae* in children (van Aartsen et al., 2019), and effects of iron supplementation in Cambodian women (Finlayson-Trick et al., 2023). However, analysis of the composition of the gut microbiota as a whole in infants from Cambodia has not previously been published.

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

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

103 Infants born into resource-poor environments can have difficulties in establishing a healthy gut
104 microbial ecosystem (Lang & MAL-ED, 2018) due to multiple factors such as birth and breast feeding
105 practices, antibiotic exposure, maternal microbiota, diet, socio-economic status, host genetics,
106 agricultural dependence, urbanisation, access to clean water and hygiene practices (Porras and
107 Brito, 2019; Robertson et al., 2019). Infant microbiomes of those in LMICs are reportedly higher in
108 *Segatella* (previously *Prevotella*) species, associated with both positive and negative health
109 outcomes, and may correlate with higher plant based, fibre rich diets (Porras and Brito 2019). The
110 phylum Pseudomonadota (previously Proteobacteria) that includes a variety of pathogenic and
111 opportunistic pathogens, is highly prevalent in microbiomes of those in LMICs (Porras and Brito
112 2019). These factors may negatively affect the composition of the early infant gut microbiota
113 (Robertson et al., 2019), and may drive these observed microbiome differences across populations
114 (Porras and Brito 2019).

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

genes. Finally, the concentrations of faecal cytokines were measured to identify any associations with the gut microbiota. We aimed to explore the composition of gut microbiota of these infants and its associations with their environment.

Results

The longitudinal infant faecal samples analysed in this study were sourced from 32 infants representing a subset of infants that are part of the NHAM birth cohort based in the rural Kampong Cham province in southeast Cambodia, and approximately 60 kilometres from the provincial capital Kampong Cham (Miyazaki et al., 2020, 2021) (**Table 1**). Samples were collected from infants living in a riverside community stretching along approximately 15.5 km of riverbank of the Mekong River from Khbop Ta Nguon to Peam Koh Sna (Miyazaki et al., 2021). The infants were living in a rural area, with most families being small-scale farmers.

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)
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)

Table 1: Infant baseline characteristics.

141 ***Sample storage conditions influence gut microbiota profiles***

142 Sample collection and storage is challenging in areas without the facilities to immediately freeze
143 faecal samples after collection. To compare the preservation of DNA in samples under different
144 conditions, half of each faecal sample was immediately placed into a DNA preservation solution
145 (DNAShield), with the remaining half remaining untreated. The faecal samples were collected in a
146 rural location with approximately 2 days between the sample being produced and final laboratory
147 storage. The untreated half of each sample was transported in iceboxes and then initially frozen at -
148 20°C and then later stored at -80°C. The half of each sample stored in DNAShield were transported
149 at ambient temperature and then stored refrigerated at 4°C. Three, biweekly samples from 32
150 infants were collected (total of 96 samples), with half split between each storage condition. Both
151 frozen and DNAShield paired infant gut microbiota samples underwent 16S rRNA amplicon
152 sequencing to analyse their bacterial composition.

153 Out of the 96 paired samples, a total of 94 were successfully sequenced when stored in DNAShield,
154 while only 70 were successfully sequenced from Frozen samples (**Supplementary File 1**). There were
155 69 pairs of samples with successful 16S rRNA amplicon sequencing from both the Frozen and
156 DNAShield portion of the same faecal sample. The resulting bacterial genus composition was
157 compared between these 69 pairs of samples (**Figure 1**). While the total number of bacterial genera
158 detected was similar between frozen and DNAShield samples (**Figure 1A**), microbiota diversity
159 calculated using the Shannon index and Inverse Simpson index was higher in DNAShield samples
160 than paired frozen samples (**Figure 1B-C**). *Bacteroides*, *Escherichia, Parabacteroides, and*
161 *Lachnoclostridium* were the four bacterial genera with the largest median difference in relative
162 abundance between storage conditions that were higher in DNAShield samples compared to frozen
163 samples (**Figure 1D-G; Supplementary File 1**). *Bacteroides* showed the largest difference in favour of
164 DNAShield storage with a median relative abundance of 0.4% in frozen samples and 8.6% in
165 DNAShield samples (**Figure 1D; Supplementary File 1**). The relative abundance of *Bifidobacterium*,
166 *Streptococcus*, *Enterococcus,* and *Actinomyces* were the genera with the largest difference and
167 significantly higher abundance in favour of frozen samples (**Figure 1H-K; Supplementary File 1**).
168 *Bifidobacterium* comprised a median relative abundance of 63% in frozen samples and 46% in
169 DNAShield samples (**Figure 1H; Supplementary File 1**). Further analysis of bacterial microbiota
170 composition in this study was performed using samples stored in DNAShield due to the larger
171 number of samples with successful sequencing results and the higher abundance of individual
172 bacterial genera surviving in DNAShield stored samples.

173 These results indicate that sample storage conditions selectively can alter the abundance of different
174 genera in analysis of bacterial composition with significantly higher abundance of *Bacteroides* and
175 *Parabacteroides* remaining when samples were stored in DNA preservation solution. This shows that
176 DNA preservation solution can preserve genus abundance in final composition analysis even when
177 samples are initially transported at ambient temperatures and then stored refrigerated and not
178 frozen.

179

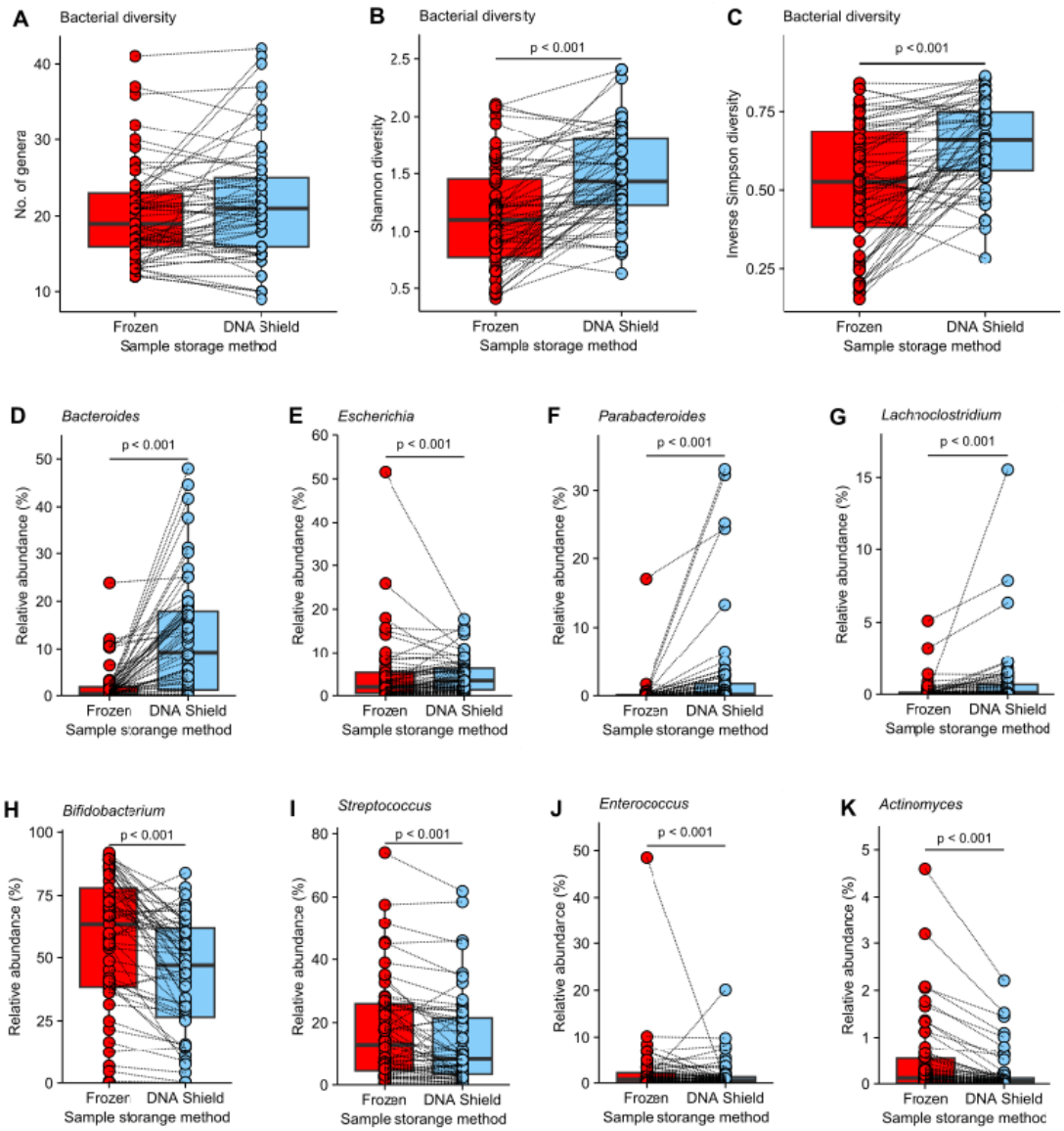


Figure 1. Differences in genus abundance between paired frozen samples and DNAShield stored samples. (A) The number of genera detected in each sample; (B) The Shannon diversity index of each sample; (C) The Inverse Simpson diversity index of each sample; Relative abundance between paired sampled after 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.

190 ***Bacterial genus composition of the infant gut microbiota is influenced by breastfeeding and***
191 ***diarrheal illness***

192 The gut microbiota of infants was dominated by *Bifidobacterium,* which comprised a median relative
193 abundance of 48.7% (**Figure 2A**). In addition, other genera including *Bacteroides, Blautia,*
194 *Erysipelatoclostridium, Escherichia, Lactobacillus, Megamonas, Parabacteroides, and Streptococcus*,
195 comprised the top ten most abundant genera, and formed the majority of the gut microbiota in
196 these infants (**Figure 2A**).

197 An exploratory analysis was carried out using metadata available for the infants from which samples
198 were collected. This included cross-sectional data collected at the start of the study before collection
199 of the faecal sample at time point 1 and additional information collected at each sample collection
200 visit (**Supplementary File 2**). The exploratory analyses used non-metric multidimensional scaling
201 (NMDS) and sample beta-diversity to cluster the samples by the similarity of their overall microbiota
202 composition at time point 1. Associations between the infant metadata with the beta-diversity of
203 the gut microbiota was analysed using Envfit and significant associations illustrated on the NMDS
204 plot (**Figure 2B**). Only breast-feeding status was significantly associated with microbiota beta-
205 diversity with clear separation in clusters between infants breast fed and those infants not breast
206 fed (**Figure 2B**).

207 To identify associations between the relative abundance of individual genera and infant metadata an
208 analysis using the MaAsLin2 package was carried out that included the samples from all three time
209 points, with the time point variable included as a random effect (**Figure 2C; Supplementary File 3**).
210 When background metadata variables were analysed individually as a fixed effect, both
211 breastfeeding status and recent diarrhoeal illness within the previous seven days showed significant
212 associations with genus abundance, which was maintained when both variables were included
213 together as a multivariable analysis (**Figure 2C**). Genera with differential abundance that was
214 significantly associated with breast-feeding and/or diarrheal illness are shown as a heat map with
215 positive and negative symbols indicating the direction of association (**Figure 2C**). The genera
216 positively associated with breast-feeding were *Bifidobacterium* and *Streptococcus* (**Figure 2C; Figure**
217 **2D**), whilst those with the strongest negative associations were *Klebsiella* and *Clostridium* (**Figure 2C**
218 **; Figure 2E**). Breast-feeding status remained unchanged for individual infants across the three time
219 points. At time point 1 the mean age of breast fed infants was 207 days and non-breast fed infants
220 214 days. Including age as a variable in the analysis did not alter the associations with breastfeeding.
221 Other measures of infant diet, the use of a feeding bottle or complementary feeding, were not
222 associated with genus abundance. MaAsLin2 analysis also showed that that recent diarrheal illness

223 within the previous seven days was associated with a lower relative abundance of *Bacteroides*
224 (**Figure 2F**) and higher relative abundance of *Clostridium*, *Klebsiella* and *Shigella* (**Figure 2G-H**).
225 *Shigella* was detected in 78% of samples, *Klebsiella* in 82% and *Clostridium* in 65% indicating that
226 these infants have a high prevalence of these potentially pathogenic genera in their gut microbiota
227 (**Supplementary File 4**). Breast feeding was the only variable associated with differences in microbial
228 diversity with the total number of bacterial genus (**Figure 3I**) and the Shannon diversity lower in
229 breast fed infants than non-breast fed infants (**Figure 3J**).

230 Only three infants included in the analysis were classified as having stunted growth with a height-
231 for-age z-score of less than 2 and this was not associated with microbiota composition. Tested as
232 continuous measures, length-for-age z-score , underweight z-score, and weight-for-length z-score
233 were also not associated with microbiota composition. However, this number of infants was too
234 small for robust testing preventing any strong conclusions being drawn from this result. Measures of
235 household hygiene facilities and total number of people in the household were not associated with
236 microbiota composition. Of particular note was the lack of an association between genus
237 composition and recent antibiotic use within the previous 7 days (**Supplementary File 3**).

238 These results indicate that breast feeding was the most influential factor shaping the gut microbiota
239 composition in favour of increased *Bifidobacterium* abundance. These results also show that recent
240 diarrheal illness was associated with higher relative abundance of the potentially pathogenic genera
241 *Clostridium*, *Klebsiella* and *Shigella*.

242

243

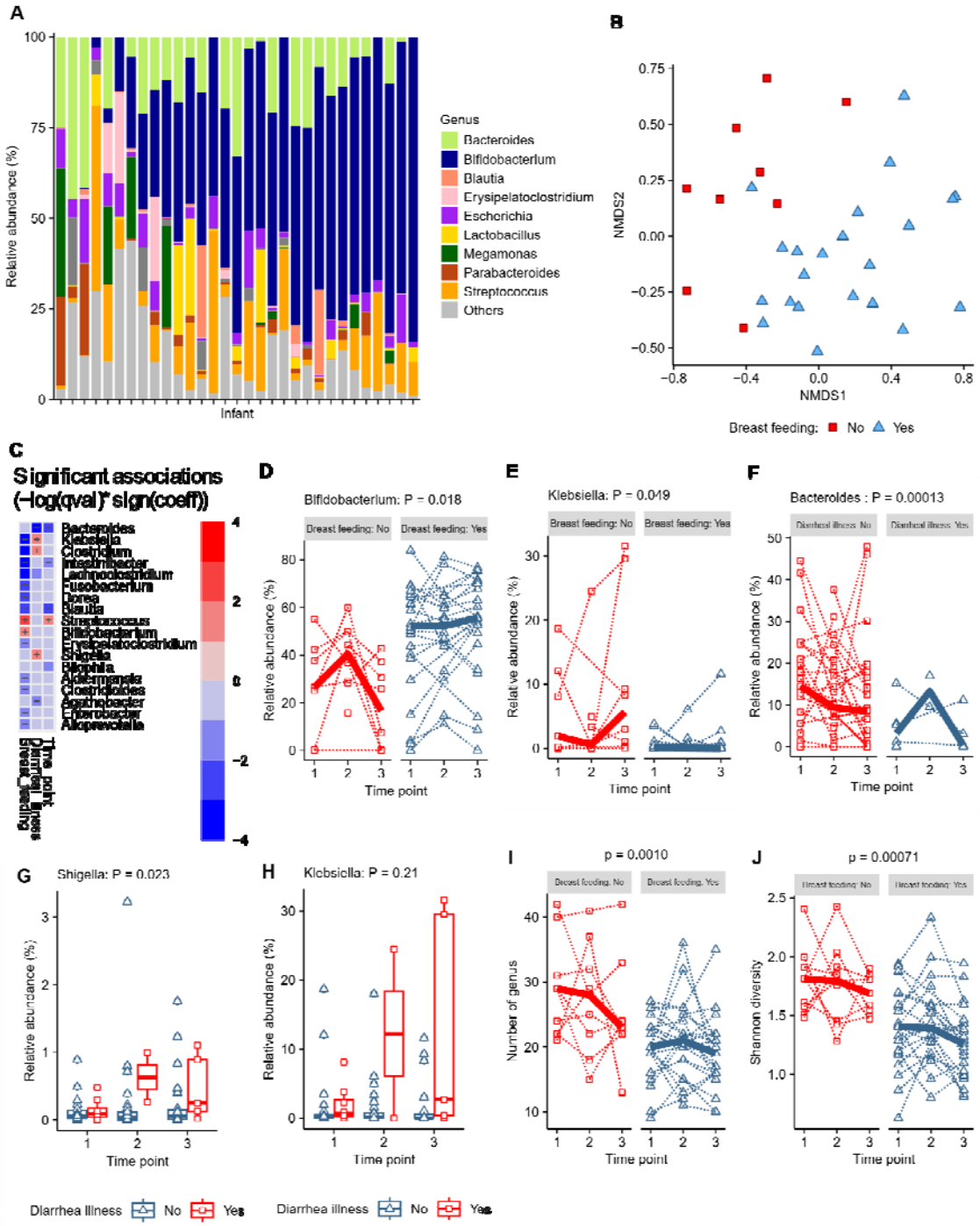


Figure 2. Genus composition and associations of genus composition with breast feeding status and diarrheal illness. (A) The ten most abundant bacterial genera across all samples ordered on the x-axis by *Bifidobacterium* abundance in each infant; (B) Non-metric multidimensional scaling (NMDS) plot representing 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

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

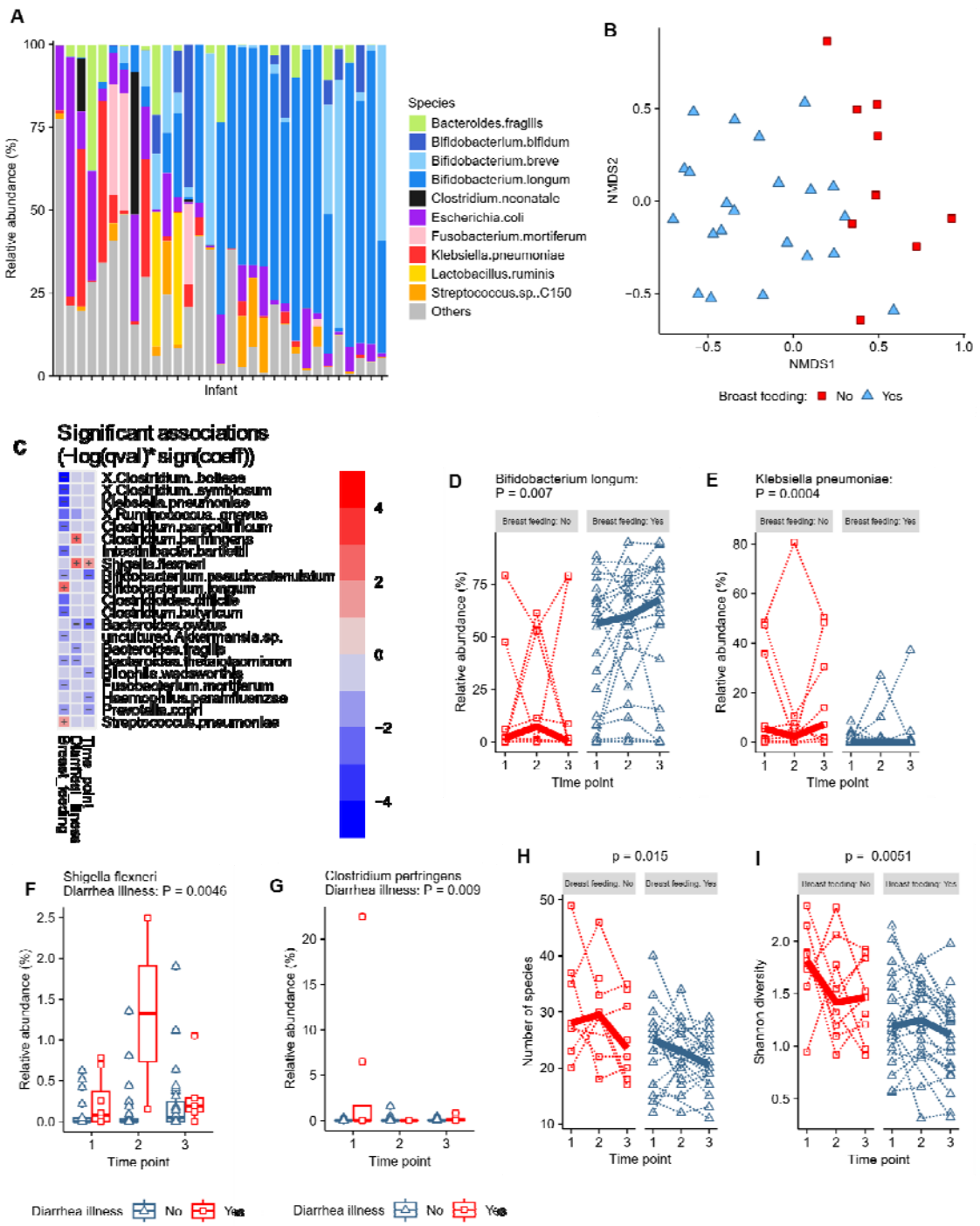
Bacterial species composition of the infant gut microbiota is influenced by breastfeeding status

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

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

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

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



295

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

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

310 ***Infants harbour bacterial strains that encode diverse human milk oligosaccharide and***
311 ***antimicrobial resistance genes***

312 The frozen faecal samples that were not stored in DNA preservation solution were used to culture
313 and isolate individual colonies for whole genome sequencing to gain further insight into the species
314 of bacteria in the gut microbiota of these infants. Genomic analysis was carried out on the
315 sequenced isolates to identify the species (and strains), and for the presence of virulence traits,
316 antimicrobial resistance genes and HMO gene clusters. From the bacterial isolates genome
317 sequenced 2 were identified as *Bacteroides*, 6 as *Bifidobacterium*, 1 as *Ruminococcus*, 3 as
318 *Enterococcus*, 21 as *Clostridium*, 3 as *Clostridioides*, and 2 as *Paraclostridium* (**Figure 4A**).

319 *Bifidobacterium* isolates comprised three species including *B. pseudocatenulatum*, *B. breve*, and *B.*
320 *longum*. Both *B. breve* and *B. longum* isolates contained complete HMO gene clusters, with none in
321 *B. pseudocatenulatum* (**Figure 4B**). This indicates the ability of the *B. breve* and *B. longum* present to
322 digest the oligosaccharides present in human breast milk, with only *B. longum* showing higher
323 relative abundance in breastfed infants compared to non-breast fed (**Figure 4C**). Comparing the four
324 genome-sequenced bacterial isolates identified as *Bifidobacterium longum* using Average Nucleotide
325 Identity with a 98% cut-off, three were identified as *B. longum* subsp. *longum* and one as *B. longum*
326 subsp. *suis* (**Supplementary File 8; Supplementary Figure 1**).

327 Genomic analysis showed the presence of antimicrobial resistance genes across all strains isolated
328 (**Figure 4D**). Of particular note was the presence of the antimicrobial resistance genes *erm*(X), *tet*(O),
329 *tet*(O/23/O), *tet*(Q), and *tet*(W) in the genomes of isolated *Bifidobacterium* species (**Figure 4B**),
330 which provide resistance to erythromycin (*erm*) and tetracycline (*tet*) respectively. The identification
331 of antimicrobial resistance genes in the *Bifidobacterium* strains present may link with the lack of any
332 association between gut microbiota composition and those infants receiving antibiotics, as shown
333 here with NMDS plots showing all samples combined (**Figure 4E**). The relative abundance of

334 *Bifidobacterium* was unchanged in samples from infants receiving antibiotics (**Figure 4F**). Amongst
335 this group of infants the most commonly received antibiotic was amoxycillin, with individual
336 instances of cefixime, and cefadroxile also received (**Supplementary File 2**).

337 Two *Clostridium perfringens* strains were isolated and genome-sequenced. Both contained five toxin
338 producing genes and six colonisation factor genes that relate to the ability to cause infection and
339 disease in the gut (**Figure 4G**). This confirms the presence of *C. perfringens* that was identified using
340 16S rRNA amplicon sequencing and the presence of toxin and colonisation factor genes supports a
341 potential role in its association with cases of infant diarrhoea.

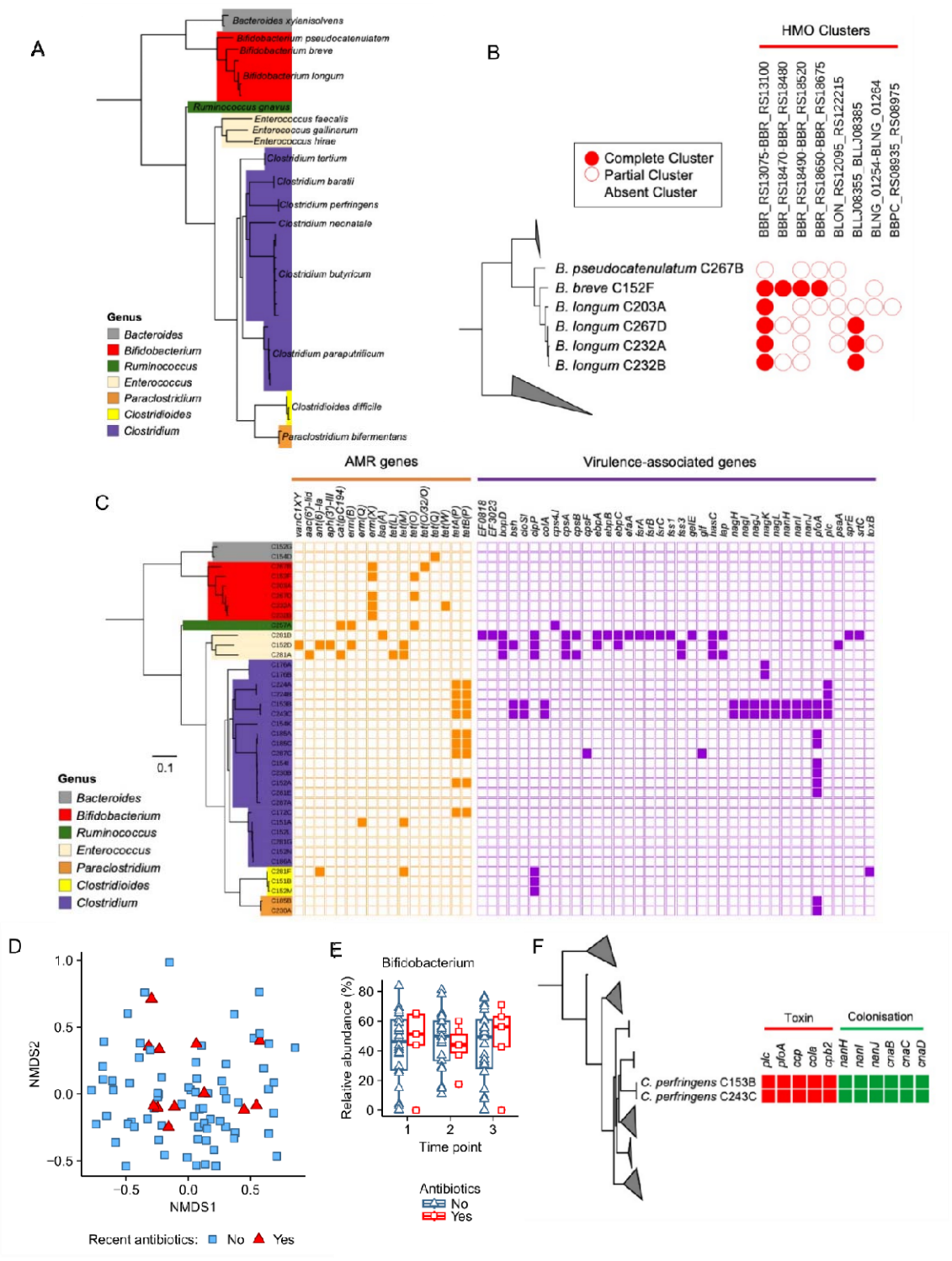


Figure 4. Genomic analysis of isolates for species determination, human milk oligosaccharide (HMO) gene clusters, antimicrobial resistance genes, and virulence genes. (A) Tree showing species and relatedness of whole genome-sequenced isolates; (B) Presence of HMO gene clusters in *Bifidobacterium* isolates;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

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

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

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

354 A linear mixed model incorporating infant metadata variables as a fixed effect and the infant as a
355 random effect was used as an exploratory analysis of associations between each cytokine and
356 metadata variables in all sample across the three time points (**Supplementary File 10**). The most
357 robust and consistent association was identified between two cytokines and breast feeding (**Figure**
358 **5A-C**). The cytokines Interleukin-1-alpha (IL-1α) (**Figure 5A**) and Vascular endothelial growth factor
359 (VEGF) (**Figure 5B**) were at significantly higher concentration in breast-fed infants. Both cytokines
360 showed substantial variation in concentration across time points within samples from the same
361 infant. IL-1α (**Figure 5A**) and VEGF (**Figure 5B**) were also the two cytokines with the highest
362 concentrations in the faecal samples across all infants.

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

371

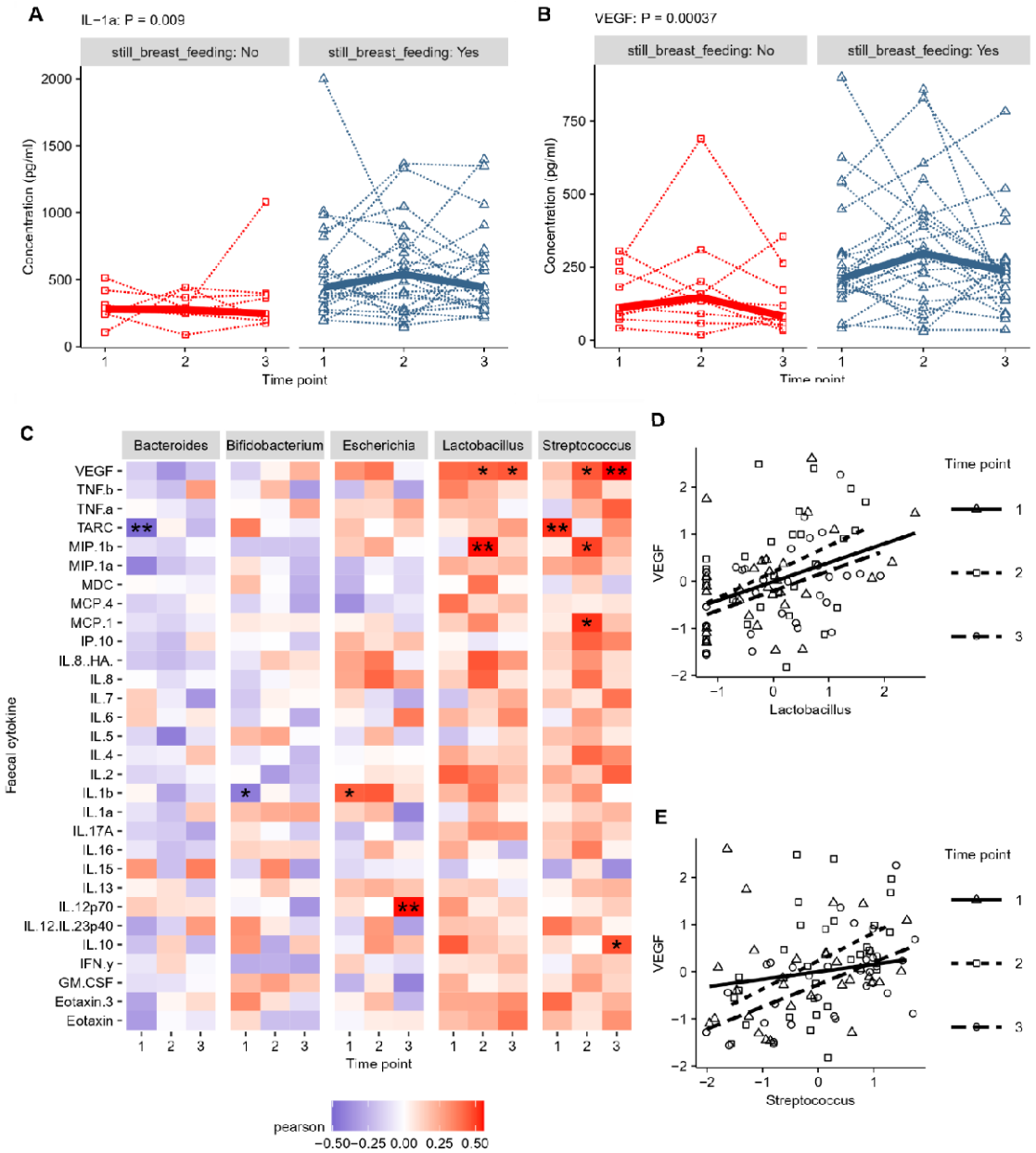


Figure 5. Infant faecal cytokines concentration and associations with breast feeding status. (A) Individual infant IL-1α concentration by breastfeeding status across time points; (B) Individual infant VEGF concentration by breastfeeding status across time points; (C) Pearson correlations between faecal 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.

381 **Discussion**

382 This study represents the first time that the infant gut microbiota composition of a cohort of rural
383 Cambodian infants has been analysed in detail. We show that shortage of samples under rural
384 conditions can radically reduce the abundance of certain genus within the samples. Infants had a gut
385 microbiota dominated by the genus *Bifidobacterium,* shaped to a large degree by breastfeeding,
386 which also impacts faecal cytokine profiles. Previous gut microbiota research in Cambodia has
387 included investigation into antimicrobial resistance genes in Enterobacterales (Auguet et al., 2021)
388 the epidemiology of antibiotic resistance in *Escherichia coli* and *Klebsiella pneumoniae* (van Aartsen
389 et al., 2019), and effects of iron supplementation on individual genus abundance in women
390 (Finlayson-Trick et al., 2023). This present study goes further and examines in-depth genus and
391 selected species and strain composition in Cambodian infants for the first time.

392 Our study results represent an unusual example of different sample storage conditions as these
393 samples were stored without immediate freezing. Overall the effect of storage conditions on the
394 composition of the microbiota was small but a higher relative abundance of *Bacteroides* in particular
395 was seen in samples stored in DNA preservation solution and a lower abundance of common species
396 including *Bifidobacterium* due to the proportional nature of the microbiota data. *Bacteroides*
397 represents an important genus in the infant gut microbiota, transferred from the mother during
398 delivery and has been found to be higher in vaginally delivered infants (Matharu et al., 2022) with
399 some *Bacteroides* species having the ability to consume components of breast milk (Kijner et al.,
400 2022). The relatively high proportion of *Bacteroides* in these infants and their almost disappearance
401 from samples without DNA preservation solution shows the importance of a storage method that
402 preserves this significant component of the infant microbiota.

403 Previous research has investigated the effects of storage conditions on microbiota analysis and
404 reported alterations in *Bacteroides* abundance, although this has focused on storage conditions
405 while frozen or number of freeze thaw cycles (Pérez-Burillo et al., 2021; Poulsen et al., 2021). Room
406 temperature storage of faecal samples for up to 72 hours has been reported to reduce genus
407 including *Bacteroides* in adult faecal samples over time (Roesch et al., 2009). While defrosting frozen
408 samples for three hours before analysis reduced *Bacteroides* abundance by half (Cardona et al.,
409 2012). The duration of cool but unfrozen temperatures that samples in the present study were
410 exposed to appear to have resulted in the particular low of *Bacteroides*. However, the unpreserved
411 samples were also important for culturing and isolating live bacteria for genomic sequencing,
412 highlighting the importance of having samples stored under both conditions.

413 *Bifidobacterium* was present at lower abundance in non-breast fed infants and not breast feeding
414 was associated with higher abundance of potentially pathogenic bacterial genus like *Klebsiella* and
415 *Clostridium*. Although the numbers of infants not breast fed was too low to draw strong conclusions,
416 previous research has indicated that increased *Bifidobacterium* in the preterm infant gut can help
417 reduce the proportions of potential pathogens (Alcon-Giner et al., 2020). This is linked to the ability
418 of *Bifidobacterium* to metabolise HMOs in breast milk and reduce gut pH to make it a more
419 inhospitable environment for overgrowth of opportunistic pathogens. *Bifidobacterium longum* was
420 the dominant species in this cohort of infants and *B. longum* was the only *Bifidobacterium* species
421 with higher abundance in breast fed infants. From the four *B. longum* isolates genome-sequenced,
422 all contained one or two HMO gene clusters indicating that they were potentially capable of
423 metabolising the sugars present in breast milk. Interestingly, three of the *B. longum* isolates were
424 confirmed to be *B. longum* subsp. *longum*, which do not typically dominate the gut microbiota of
425 infants in developed western countries nor other LMICs with high rates of breast feeding. It was
426 recently reported that *B. longum* subsp. *infantis* was the dominant *Bifidobacterium* in 1 to 2 month
427 old infants from the Gambia and Bangladesh with *B. longum* subsp. *longum* only forming a minor
428 proportion of the microbiota (Taft et al., 2022). The isolation of *B. longum* subsp. *longum* from
429 Cambodian samples may be due to sampling at an older age of around 7 months, with recent
430 research from Bangladesh indicating that a transition occurs from *B. longum* subsp. *infantis* to *B.*
431 *longum* subsp. *longum* with increasing infant age (Vatanen et al., 2022). One *B. longum* isolate was
432 identified as most closely matching *B. longum* subsp. *suis*, a subspecies originally isolated from pigs
433 (Matteuzzi et al., 1971) and found to be a dominant member of the piglet gut microbiota (Zani et al.,
434 1974). *Bifidobacterium* isolates corresponding to *B. longum* subsp. *suis* were recently identified in
435 Bangladeshi infants (Vatanen et al., 2022), which suggests that this may be an under-recognised
436 member of the infant gut microbiota. From only a single isolate we cannot determine the proportion
437 of *B. longum* subsp. *suis* within Cambodian samples, and this could represent the transient
438 colonisation of this bacterium from close proximity to pigs or that this is a resident of the gut
439 microbiota in this region. The presence of an HMO gene cluster capable of metabolising breast milk
440 sugars indicates that this *B. longum* subsp. *suis* isolate may be capable of metabolising breast milk
441 HMOs. The dominant abundance of the species *B. longum* in breastfed infants underline the central
442 effect of diet on the gut microbiota composition of these infants, overwhelming any other
443 environmental factor in this small cohort.

444 Previous research has identified associations between restricted infant growth and stunting due to
445 malnutrition and differences in the infant gut microbiota (Monira et al., 2011; Surono et al., 2021).
446 We did not find any associations between the gut microbiota and infants growth. This was

447 potentially due to the small cohort of infants involved with only three out of the thirty infants
448 included showing stunted growth and the relatively young age of the infants starting at an average
449 of 7 months of age. Previous research has looked at older children (Monira et al., 2011) or older
450 infants with a higher proportion of the infants suffering from stunting (Surono et al., 2021). This
451 indicates that in this small cohort of infants it is the presence or absence of breast milk having the
452 largest effect, potentially overwhelming the ability to detect other environmental influences.

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

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

481 *Bifidobacterium* abundance despite antibiotic treatment may point to the presence of other novel
482 resistance genes in these *Bifidobacterium*, which if identified may have future potential for
483 combining effective probiotic therapy with antibiotic treatment.

484 One genus notable by its absence from this cohort of Cambodian infants is the genus *Segatella*.
485 Previously found to be a dominant member of the gut microbiota in rural West Africa while rare in
486 Italian children (De Filippo et al., 2010). More recently this high prevalence of *Segatella* in children in
487 low income countries has been confirmed in The Gambia, Kenya, Mali, and Bangladesh (Pop et al.,
488 2014). Whilst *Segatella* abundance was lower in children from Bangladesh it was still detectable in
489 children of a similar age to this cohort of Cambodian infants (Pop et al., 2014). In contrast
490 Cambodian infants contained *Bacteroides*, a genus previously associated with the gut microbiota of
491 European and US inhabitants (De Filippo et al., 2010). What dietary differences may account for this
492 difference cannot be determined in this study, however, it does show that further investigation is
493 needed into underrepresented populations. Whether the presence of *Bacteroides* and absence of
494 *Segatella* persists into adulthood remains unknown as composition of the adult gut microbiota in
495 this region remains still to be investigated.

496 Consistent differences in the faecal concentrations of the cytokines IL-1α and VEGF point to
497 potential effects of breastfeeding on the infant gut mucosal immune system. VEGF has been
498 reported to be secreted by the mammary glands, and breast milk contains high concentrations of
499 free VEGF that can bind to receptors on the surface of cells in the infant intestine (Kobata et al.,
500 2008; Siafakas et al., 1999; Vuorela et al., 2000). Reduced VEGF in preterm infants is suspected to be
501 involved in necrotising enterocolitis which is associated with overgrowth of opportunistic pathogens
502 including *C. perfringens* (Kiu et al., 2023; Sabnis et al., 2015). While the role of VEGF originating from
503 breast milk in term infants remains unexplored this growth factors could exert growth-promoting
504 and protective effects on the infant gut (Kobata et al., 2008). While the role of IL-1α in the infant gut
505 has not been explored, it has also been reported to be present in breast milk (Zanardo et al., 2005)
506 however, whether this has any effects the infant gut is currently unknown. Although we are unable
507 to demonstrate a breast milk origin from these results, the presence of higher concentrations of
508 these three faecal cytokines in breast fed infants does suggest that they may be derived from breast
509 milk, with a potential to influence infant development that warrants further investigation.

510 While there was little consistent association across time points between the gut microbiota at the
511 genus level and faecal cytokines, a weak positive correlation was noted between *Lactobacillus* and
512 *Streptococcus* and the faecal cytokine VEGF. As *Streptococcus* abundance was higher in breastfed
513 infants this correlation is potentially linked to breastfeeding. As both microbiota composition and

514 faecal cytokines is variable with time it may not be possible to identify links between factors in a
515 limited sample size of infants. Interestingly, despite being the genus most differentiated by breast
516 feeding status there no consistent correlations between *Bifidobacterium* IL-1α or VEGF. Indeed,
517 many species and strains from this genus are known to positively influence the gut and systemic
518 immune system – including *B. longum,* which is proposed to be in part mediated by metabolites
519 produced as a result of HMO degradation e.g. acetate (Henrick et al., 2021; Laursen et al., 2021).
520 However, further studies including larger number of infants would need to be performed to explore
521 microbial-immune interactions in more detail. In addition, the use of the frozen samples for faecal
522 cytokine analysis may be limited as it is unknown how such storage conditions may impact the
523 degradation of the cytokines tested.

524 We show that this cohort of infants from rural Cambodia have a *Bifidobacterium* dominated gut
525 microbiota (including atypical species) that is primarily shaped by breastfeeding. This highlights the
526 influence of breast feeding in supporting a beneficial infant microbiota in rural conditions. As
527 *Bacteroides* represents a common and important genus in the human gut the effects of sample
528 storage are an important consideration when immediate freezing of samples is not possible. The lack
529 of impact of antibiotic use may indicates either widespread antibiotic resistance or the use of
530 ineffective antibiotics. The finding that faecal cytokines are shaped by breast feeding and could
531 potentially originate from breast milk is a finding with unknown but potentially interesting
532 implications. This study adds significantly to the available data on the composition of infant gut
533 microbiota communities in a previously neglected and difficult to sample geographical area of the
534 world. Moreover, given Cambodian profiles are somewhat different from other infant cohorts,
535 including LMICs, studies such as these shows the importance of looking at infants from
536 underrepresented areas for defining what can constitute a normal infant microbiota.

537

538

539

540

541

542

543

544

Materials & Methods

Study design & ethics

Samples were collected from a sub-group of infants enrolled in NHAM cohort based in villages along the Mekong river in rural Kampong Cham province in Cambodia. This registered all infants born between April 2016 and March 2019 with the objective of investigating causes of stunting. Full ethical approvals including National Ethical Committee for Health Research in the Ministry of Health, Cambodia (016) and the Ethical Committee of the National Centre for Global Health and Medicine in Japan (NCGM-G-001870-01), and methods for the collection of detailed meta-data for infants included in this sub-study are as previously described (Miyazaki et al., 2021, 2020). Weight and length were measured to the nearest 0.1kg and 1 mm by trained field workers at each home visit using a weighing scale (Seca 877, Seca, Hamburg, Germany) and a length board (Seca 417, Seca).

Sample collection

Infant faecal samples were analysed from a subset of 32 infants out of 47 enrolled in a detailed 3-month longitudinal follow-up study (Miyazaki et al., 2021, 2020) collected between March and July in 2017. Mothers/primary care-givers and their infants were visited by study field workers every 14 days. Three samples were analysed from each infant for a total of 96 individual samples. The trained field worker gave instructions on how to collect and equipped the caregivers with the material for the child stool collection prior to routine home visits. Once stool sample collection by mothers or caregivers was completed, the sample was kept in relatively cooler place in the house or in cooler bags with ice until the field worker visited to collect them. The majority of the samples were collected by the field worker within 10 hours of the time of child defecation reported by mothers (maximum: 47 hours and minimum: 15 min among all sample). Each sample collected was divided in two and half stored in DNAShield and stored unaltered at ambient temperatures until delivered to laboratory. The other half of the sample was stored in iceboxes with refrigerant and transported to a -20° freezer on the same day and once delivered was stored at -80°. The first sample was collected after an initial visit where cross-sectional metadata was collected. Subsequent samples were collected at approximately two week intervals along with further individual visit metadata. Underweight z-scores (weight-for-age), weight-for-length z-scores, length-for-age z-score were calculated previously (Miyazaki et al., 2021). A hand washing score was generated as a composite score combining the frequency and consistency of handwashing with soap. Improved toilet facility was defined as having access to a toilet flushing to a septic tank or pit latrine compared to unimproved having only access to a pit latrine or no toilet facility. Household toilet facility was defined as having a toilet within the home flushing to a septic tank.

578 ***DNA extraction of preterm stool samples***

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

582 ***16S rRNA amplicon sequencing of faecal samples***

583 16S rRNA region (V1-V2) primers were used for library construction as described previously (Alcon-
584 Giner et al., 2020). This set of primers allowed the amplification of one 16S rRNA gene sequencing
585 library containing 96 different samples. PCR conditions used were: cycle of 94°C 3 min and 25 cycles
586 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
587 performed using Illumina MiSeq platform with 300 bp paired end reads.

588 ***16S rRNA amplicon data analysis***

589 The raw sequence reads obtained were subjected to quality filtering using trim_galore v0.4.3. The
590 resulting filtered sequence reads were then mapped against the 16S rRNA SILVA database
591 (SILVA_132_SSURef_tax_silva) using BLASTn v2.2.25+ with a maximum e-value threshold of 10e-3.
592 This mapping process was performed separately for both pairs of sequences (Quast et al., 2013).
593 After completing the BLASTn alignments, the output files were annotated using the paired-end
594 protocol of MEGAN6, employing the default Lowest Common Ancestor (LCA) parameters (Huson et
595 al., 2016; Huson and Mitra, 2012).

596 Statistical analysis was conducted using R version 4.2.2 (2022-10-31 ucrt) and RStudio
597 (2023.06.0+421). To determine the number of reads required to adequately represent the
598 microbiota in each sample, rarefaction curves were generated using the vegan package in R. For the
599 analysis at the genus level, the sample with the smallest number of reads contained 37,306 classified
600 reads, which was deemed an acceptable minimum. For the species level analysis, the sample with
601 the second lowest number of reads contained 9,317 reads, which was considered an acceptable
602 minimum. The 16S rRNA gene sequence data were then normalized by subsampling using the rarefy
603 function in the phyloseq package, resulting in an even depth of 37,306 reads for the genus data and
604 9,317 reads for the species data. One sample was excluded from the species data analysis as it only
605 contained 3,965 reads.

606 To examine differences in the relative abundance of genera among the 69 infant fecal samples,
607 which had 16S rRNA sequence data obtained from both samples stored in DNAShield solution and
608 Frozen, a Wilcoxon rank sum test was performed in R. This test was chosen due to the non-
609 parametric nature of the majority of the genus abundance data. To account for multiple testing, a

610 Benjamini & Hochberg false discovery rate (FDR) method was applied, and a p-value of less than 0.05
611 was considered statistically significant for adjusted p-values.

612 The overall composition of the 16S rRNA gene sequence data was analyzed using NMDS (Non-metric
613 multidimensional scaling) plots. These plots were generated in R Studio using the vegan package
614 version 2.6-4, employing a Bray-Curtis dissimilarity calculation. To determine the statistical
615 significance of differences in microbial community structure based on infant metadata variables,
616 permutational MANOVA was conducted using the Adonis function from the vegan R package version
617 2.6-4.

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

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

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

636 Anaerobic bacterial culturing and isolation was performed on seventeen Cambodian infant faecal
637 samples (C151, C152, C153, C154, C172, C176, C185, C186, C203, C224, C230, C232, C243, C257,
638 C267, C281, and C287) following (Browne, 2018; Browne et al., 2016; Forster et al., 2019). Sample
639 preparation was performed under anaerobic conditions. YCFA media and PBS was pre-reduced in an
640 anaerobic cabinet at 37 ˚C for a minimum of 12hs before culturing. Two samples, weighing 100mg
641 each, were prepared from each faecal sample. The first sample was homogenised in reduced PBS,

642 serially diluted, and plated directly onto the YCFA agar. The second sample was ethanol treated
643 using 70% ethanol under ambient aerobic conditions for 1h to kill vegetative cells; the sample was
644 then washed three times with reduced PBS to remove the ethanol, homogenised in the reduced PBS,
645 serially diluted and plated onto YCFA + Sodium taurocholate agar to stimulate spore germination. All
646 agar plates were placed in the anaerobic cabinet at 37˚C for 48hs. Single colonies were picked and
647 streaked onto pre-reduced YCFA media to isolate and purify the colonies; YCFA media was pre-
648 reduced in the anaerobic cabinet at 37 ˚C for a minimum of 12hs before isolating colonies. Colonies
649 from the non-ethanol treated samples were streaked onto the reduced YCFA agar plates, and
650 colonies from the ethanol treated samples were streaked onto the reduced YCFA + Sodium
651 taurocholate agar plates. Plates were then left to incubate in the anaerobic cabinet at 37˚C for 48-
652 72hs. This was carried out a second time to make sure isolates were pure. Falcon tubes containing
653 reduced YCFA broth was inoculated with a single colony from the YCFA and YCFA + Sodium
654 taurocholate agar plates, and left to incubate in the anaerobic cabinet at 37˚C for 48hs, with
655 bacterial stocks prepared in 1.5ml cryovials with 30% glycerol YCFA broth, and stored at -80˚C.
656 Additionally, two infant faecal samples C203 and C232 were processed for targeted *Bifidobacterium*
657 isolation. Briefly, serially diluted aliquots of faecal homogenates (10^{-1} – 10^{-6} in PBS) were plated onto
658 MRS agar supplemented with 50mg/l mupirocin and 0.5g/l L-cysteine HCL and incubated in an
659 anaerobic cabinet at 37 °C for 48h. Colonies with desired morphology were randomly selected and
660 re-streaked on MRS agar to purity. Pure cultures were prepared in 1.5 ml cryovials with 30% glycerol
661 Reinforced Clostridial Medium (RCM) and stored at –80 °C.

662 ***Whole Genome Sequencing (WGS)***

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

668 ***Genomic analysis***

669 Paired-end raw sequence reads (fastq) were quality-filtered (-q 20) via fastp v0.20.0 (Chen et al.,
670 2018) prior to de novo genome assembly using SPAdes v3.14.1 (Bankevich et al., 2012) at default
671 parameters. Next, contigs smaller than 500bp were discarded in each genome assembly. All genome
672 assemblies were subjected to contamination check via CheckM v1.1.3 (Parks et al., 2015),
673 contaminated (contamination >5%) and/or incomplete (completeness <90%) genome assemblies

674 were excluded from further analyses (n=5). All genomes were identified taxonomically using gtdb-tk
675 v1.5.1 (Chaumeil et al., 2020), using ANI cut-off 95% against type strains of each species.

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

680 These genomes were also subjected to sequence search analysis via ABRicate v1.0.1 (Seemann,
681 2023) based on ResFinder v4.0 database (Bortolaia et al., 2020) (--minid=95 and --mincov=90) to
682 determine antimicrobial resistance (AMR) profiles, and VFDB database (Liu et al., 2019) (--minid=90
683 and –mincov=90) to determine virulence profiles. Furthermore, HMO clusters were detected using
684 an in-house HMO sequence database via ARBricate at –minid=70 (Lawson et al., 2020). In addition,
685 *C. perfringens* virulence profiles were predicted via database TOXIper v1.1 using ABRicate at –
686 minid=90 (Kiu, 2023).

687 For *Bifidobacterium longum* subspecies analysis, *B. longum* strains were compared with four
688 subspecies type strains namely subspecies *infantis*, subspecies *longum*, subspecies *suillum* and
689 subspecies *suis*, to generate a core gene alignment via Panaroo v1.2.8 (Tonkin-Hill et al., 2020),
690 resulting in 1,212 core genes, which core-gene alignment constructed by MAFFT (Katoh and
691 Standley, 2013) were SNP-extracted (via SNP-sites v2.3.3)(Page et al., 2016) and subsequently used
692 to generate a phylogenetic tree via IQ-TREE v2.0.5 (Minh et al., 2020, p. 2) with ultra-fast bootstrap
693 replicates -B 1000 and automatic evolutionary model selection option -m TEST (best fit model was
694 determined to be GTR+F+ASC+G4). Subspecies identification was performed using fastANI v1.3 (Jain
695 et al., 2018) at cut-off 98%.

696 ***Faecal cytokine measurement***

697 Faecal samples were homogenized with PBS using a FastPrep® Bead Beater (4.0m/s, 3min),
698 centrifuged (14,000rpm, 15min) and 25µl of supernatant was used for the assay. Samples were
699 analysed using MULTI-SPOT™ plates, MESO Quickplex SQ120 and discovery workbench software
700 according to the manufacturer's protocol. Pre-coated immunoassays V-PLEX Proinflammatory Panel
701 1 (human) and V-PLEX Cytokine Panel 1 (human) were used to detect a set of 20 different cytokines:
702 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-
703 15, IL-16, IL-17A, TNFβ, and VEGF-A.

704 ***Data and Code Availability***

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

708 ***Statistical analysis***

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

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

723

724 **Author contributions**

725 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.,
726 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.;
727 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.,
728 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.,
729 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 -
730 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
731 L.J.H; Visualisation, M.J.D and R.K.; Supervision, S.E.C,. and L.J.H.; Project Administration, M.J.D.,
732 S.E.C, and L.J.H.; Funding Acquisition, S.E.C, and L.J.H.

733

734 **Competing interest statement**

735 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., Swann, J.R., Kroll, J.S., Clarke, P., Hall, L.J., 2020. Microbiota Supplementation with Bifidobacterium and Lactobacillus Modifies the Preterm Infant Gut Microbiota and Metabolome: An Observational Study. Cell Rep Med 1, 100077. https://doi.org/10.1016/j.xcrm.2020.100077

Auguet, O.T., Niehus, R., Gweon, H.S., Berkley, J.A., Waichungo, J., Njim, T., Edgeworth, J.D., Batra, R., Chau, K., Swann, J., Walker, S.A., Peto, T.E.A., Crook, D.W., Lamble, S., Turner, P., Cooper, B.S., Stoesser, N., 2021. Population-level faecal metagenomic profiling as a tool to predict antimicrobial resistance in Enterobacterales isolates causing invasive infections: An exploratory study across Cambodia, Kenya, and the UK. EClinicalMedicine 36, 100910. https://doi.org/10.1016/j.eclinm.2021.100910

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19, 455–477. https://doi.org/10.1089/cmb.2012.0021

766 Bortolaia, V., Kaas, R.S., Ruppe, E., Roberts, M.C., Schwarz, S., Cattoir, V., Philippon, A., Allesoe, R.L.,
767 Rebelo, A.R., Florensa, A.F., Fagelhauer, L., Chakraborty, T., Neumann, B., Werner, G.,
768 Bender, J.K., Stingl, K., Nguyen, M., Coppens, J., Xavier, B.B., Malhotra-Kumar, S., Westh, H.,
769 Pinholt, M., Anjum, M.F., Duggett, N.A., Kempf, I., Nykäsenoja, S., Olkkola, S., Wieczorek, K.,
770 Amaro, A., Clemente, L., Mossong, J., Losch, S., Ragimbeau, C., Lund, O., Aarestrup, F.M.,
771 2020. ResFinder 4.0 for predictions of phenotypes from genotypes. Journal of Antimicrobial
772 Chemotherapy 75, 3491–3500. https://doi.org/10.1093/jac/dkaa345
773 Browne, H.P., 2018. Development of anaerobic culturing methods to study the evolution and
774 transmission of the human intestinal microbiota (doctoral). London School of Hygiene &
775 Tropical Medicine. https://doi.org/10.17037/PUBS.04647551
776 Browne, H.P., Forster, S.C., Anonye, B.O., Kumar, N., Neville, B.A., Stares, M.D., Goulding, D., Lawley,
777 T.D., 2016. Culturing of "unculturable" human microbiota reveals novel taxa and extensive
778 sporulation. Nature 533, 543–546. https://doi.org/10.1038/nature17645
779 Browne, H.P., Iqbal, N.T., Osman, M., Tigoi, C., Lawley, T.D., Gordon, J.I., Ahmed, T., Kariuki, S., 2024.
780 Boosting microbiome science worldwide could save millions of children's lives. Nature 625,
781 237–240. https://doi.org/10.1038/d41586-024-00017-8
782 Cambodia Demographic and Health Survey, 2022. Cambodia Demographic and Health Survey [WWW
783 Document]. URL https://dhsprogram.com/publications/publication-PR136-Preliminary-
784 Reports-Key-Indicators-Reports.cfm (accessed 1.19.23).
785 Cardona, S., Eck, A., Cassellas, M., Gallart, M., Alastrue, C., Dore, J., Azpiroz, F., Roca, J., Guarner, F.,
786 Manichanh, C., 2012. Storage conditions of intestinal microbiota matter in metagenomic
787 analysis. BMC Microbiology 12, 158. https://doi.org/10.1186/1471-2180-12-158
788 Chaumeil, P.-A., Mussig, A.J., Hugenholtz, P., Parks, D.H., 2020. GTDB-Tk: a toolkit to classify
789 genomes with the Genome Taxonomy Database. Bioinformatics 36, 1925–1927.
790 https://doi.org/10.1093/bioinformatics/btz848
791 Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor.
792 Bioinformatics 34, i884–i890. https://doi.org/10.1093/bioinformatics/bty560
793 Chung The, H., Nguyen Ngoc Minh, C., Tran Thi Hong, C., Nguyen Thi Nguyen, T., Pike, L.J., Zellmer,
794 C., Pham Duc, T., Tran, T.-A., Ha Thanh, T., Van, M.P., Thwaites, G.E., Rabaa, M.A., Hall, L.J.,
795 Baker, S., 2021. Exploring the Genomic Diversity and Antimicrobial Susceptibility of
796 Bifidobacterium pseudocatenulatum in a Vietnamese Population. Microbiology Spectrum 9,
797 e00526-21. https://doi.org/10.1128/Spectrum.00526-21
798 De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J.B., Massart, S., Collini, S.,
799 Pieraccini, G., Lionetti, P., 2010. Impact of diet in shaping gut microbiota revealed by a

800 comparative study in children from Europe and rural Africa. Proc. Natl. Acad. Sci. U.S.A. 107,
801 14691–14696. https://doi.org/10.1073/pnas.1005963107
802 Finlayson-Trick, E., Nearing, J., Fischer, J.A., Ma, Y., Wang, S., Krouen, H., Goldfarb, D.M., Karakochuk,
803 C.D., 2023. The Effect of Oral Iron Supplementation on Gut Microbial Composition: a
804 Secondary Analysis of a Double-Blind, Randomized Controlled Trial among Cambodian
805 Women of Reproductive Age. Microbiol Spectr 11, e0527322.
806 https://doi.org/10.1128/spectrum.05273-22
807 Forster, S.C., Kumar, N., Anonye, B.O., Almeida, A., Viciani, E., Stares, M.D., Dunn, M., Mkandawire,
808 T.T., Zhu, A., Shao, Y., Pike, L.J., Louie, T., Browne, H.P., Mitchell, A.L., Neville, B.A., Finn, R.D.,
809 Lawley, T.D., 2019. A human gut bacterial genome and culture collection for improved
810 metagenomic analyses. Nat Biotechnol 37, 186–192. https://doi.org/10.1038/s41587-018-
811 0009-7
812 Harris, S.R., Feil, E.J., Holden, M.T.G., Quail, M.A., Nickerson, E.K., Chantratita, N., Gardete, S.,
813 Tavares, A., Day, N., Lindsay, J.A., Edgeworth, J.D., de Lencastre, H., Parkhill, J., Peacock, S.J.,
814 Bentley, S.D., 2010. Evolution of MRSA during hospital transmission and intercontinental
815 spread. Science 327, 469–474. https://doi.org/10.1126/science.1182395
816 Henrick, B.M., Rodriguez, L., Lakshmikanth, T., Pou, C., Henckel, E., Arzoomand, A., Olin, A., Wang, J.,
817 Mikes, J., Tan, Z., Chen, Y., Ehrlich, A.M., Bernhardsson, A.K., Mugabo, C.H., Ambrosiani, Y.,
818 Gustafsson, A., Chew, S., Brown, H.K., Prambs, J., Bohlin, K., Mitchell, R.D., Underwood,
819 M.A., Smilowitz, J.T., German, J.B., Frese, S.A., Brodin, P., 2021. Bifidobacteria-mediated
820 immune system imprinting early in life. Cell 184, 3884-3898.e11.
821 https://doi.org/10.1016/j.cell.2021.05.030
822 Huson, D.H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., Ruscheweyh, H.-J., Tappu, R.,
823 2016. MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale
824 Microbiome Sequencing Data. PLOS Computational Biology 12, e1004957.
825 https://doi.org/10.1371/journal.pcbi.1004957
826 Huson, D.H., Mitra, S., 2012. Introduction to the Analysis of Environmental Sequences:
827 Metagenomics with MEGAN, in: Anisimova, M. (Ed.), Evolutionary Genomics: Statistical and
828 Computational Methods, Volume 2, Methods in Molecular Biology. Humana Press, Totowa,
829 NJ, pp. 415–429. https://doi.org/10.1007/978-1-61779-585-5_17
830 Jain, C., Rodriguez-R, L.M., Phillippy, A.M., Konstantinidis, K.T., Aluru, S., 2018. High throughput ANI
831 analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9, 5114.
832 https://doi.org/10.1038/s41467-018-07641-9

833 Katoh, K., Standley, D.M., 2013. MAFFT Multiple Sequence Alignment Software Version 7:
834 Improvements in Performance and Usability. Molecular Biology and Evolution 30, 772–780.
835 https://doi.org/10.1093/molbev/mst010
836 Katz, L.S., Griswold, T., Morrison, S.S., Caravas, J.A., Zhang, S., den Bakker, H.C., Deng, X., Carleton,
837 H.A., 2019. Mashtree: a rapid comparison of whole genome sequence files. J Open Source
838 Softw 4. https://doi.org/10.21105/joss.01762
839 Kijner, S., Cher, A., Yassour, M., 2022. The Infant Gut Commensal Bacteroides dorei Presents a
840 Generalized Transcriptional Response to Various Human Milk Oligosaccharides. Frontiers in
841 Cellular and Infection Microbiology 12.
842 Kiu, R., 2023. TOXIper: Toxinotype assignment of Clostridium perfringens via ABRicate.
843 Kiu, R., Shaw, A.G., Sim, K., Acuna-Gonzalez, A., Price, C.A., Bedwell, H., Dreger, S.A., Fowler, W.J.,
844 Cornwell, E., Pickard, D., Belteki, G., Malsom, J., Phillips, S., Young, G.R., Schofield, Z., Alcon-
845 Giner, C., Berrington, J.E., Stewart, C.J., Dougan, G., Clarke, P., Douce, G., Robinson, S.D.,
846 Kroll, J.S., Hall, L.J., 2023. Particular genomic and virulence traits associated with preterm
847 infant-derived toxigenic Clostridium perfringens strains. Nat Microbiol 8, 1160–1175.
848 https://doi.org/10.1038/s41564-023-01385-z
849 Kobata, R., Tsukahara, H., Ohshima, Y., Ohta, N., Tokuriki, S., Tamura, S., Mayumi, M., 2008. High
850 levels of growth factors in human breast milk. Early Human Development 84, 67–69.
851 https://doi.org/10.1016/j.earlhumdev.2007.07.005
852 Lackey, K.A., Williams, J.E., Meehan, C.L., Zachek, J.A., Benda, E.D., Price, W.J., Foster, J.A., Sellen,
853 D.W., Kamau-Mbuthia, E.W., Kamundia, E.W., Mbugua, S., Moore, S.E., Prentice, A.M., K.,
854 D.G., Kvist, L.J., Otoo, G.E., García-Carral, C., Jiménez, E., Ruiz, L., Rodríguez, J.M., Pareja,
855 R.G., Bode, L., McGuire, M.A., McGuire, M.K., 2019. What's Normal? Microbiomes in Human
856 Milk and Infant Feces Are Related to Each Other but Vary Geographically: The INSPIRE Study.
857 Front. Nutr. 6, 45. https://doi.org/10.3389/fnut.2019.00045
858 Laursen, M.F., Sakanaka, M., von Burg, N., Mörbe, U., Andersen, D., Moll, J.M., Pekmez, C.T.,
859 Rivollier, A., Michaelsen, K.F., Mølgaard, C., Lind, M.V., Dragsted, L.O., Katayama, T.,
860 Frandsen, H.L., Vinggaard, A.M., Bahl, M.I., Brix, S., Agace, W., Licht, T.R., Roager, H.M.,
861 2021. Bifidobacterium species associated with breastfeeding produce aromatic lactic acids in
862 the infant gut. Nat Microbiol 6, 1367–1382. https://doi.org/10.1038/s41564-021-00970-4
863 Lawson, M.A.E., O'Neill, I.J., Kujawska, M., Gowrinadh Javvadi, S., Wijeyesekera, A., Flegg, Z.,
864 Chalklen, L., Hall, L.J., 2020. Breast milk-derived human milk oligosaccharides promote
865 Bifidobacterium interactions within a single ecosystem. ISME J 14, 635–648.
866 https://doi.org/10.1038/s41396-019-0553-2

867 Letunic, I., Bork, P., 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments.
868 Nucleic Acids Research 47, W256–W259. https://doi.org/10.1093/nar/gkz239
869 Liu, B., Zheng, D., Jin, Q., Chen, L., Yang, J., 2019. VFDB 2019: a comparative pathogenomic platform
870 with an interactive web interface. Nucleic Acids Research 47, D687–D692.
871 https://doi.org/10.1093/nar/gky1080
872 Mallick, H., Rahnavard, A., McIver, L.J., Ma, S., Zhang, Y., Nguyen, L.H., Tickle, T.L., Weingart, G., Ren,
873 B., Schwager, E.H., Chatterjee, S., Thompson, K.N., Wilkinson, J.E., Subramanian, A., Lu, Y.,
874 Waldron, L., Paulson, J.N., Franzosa, E.A., Bravo, H.C., Huttenhower, C., 2021. Multivariable
875 association discovery in population-scale meta-omics studies. PLOS Computational Biology
876 17, e1009442. https://doi.org/10.1371/journal.pcbi.1009442
877 Matharu, D., Ponsero, A.J., Dikareva, E., Korpela, K., Kolho, K.-L., de Vos, W.M., Salonen, A., 2022.
878 Bacteroides abundance drives birth mode dependent infant gut microbiota developmental
879 trajectories. Frontiers in Microbiology 13.
880 Matteuzzi, D., Crociani, F., Zani, O., Trovatelli, L.D., 1971. Bifidobacterium suis n. sp.: A new species
881 of the genus Bifidobacterium isolated from pig faces. Zeitschrift für allgemeine Mikrobiologie
882 11, 387–395. https://doi.org/10.1002/jobm.19710110504
883 Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear,
884 R., 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the
885 Genomic Era. Molecular Biology and Evolution 37, 1530–1534.
886 https://doi.org/10.1093/molbev/msaa015
887 Miyazaki, A., Matsui, M., Tung, R., Taing, B., White, L.V., Iwamoto, A., Cox, S.E., 2021. Determinants
888 of growth measurements in rural Cambodian infants: a cross-sectional study. Int Health 13,
889 49–56. https://doi.org/10.1093/inthealth/ihaa018
890 Miyazaki, A., Tung, R., Taing, B., Matsui, M., Iwamoto, A., Cox, S.E., 2020. Frequent unregulated use
891 of antibiotics in rural Cambodian infants. Transactions of The Royal Society of Tropical
892 Medicine and Hygiene 114, 401–407. https://doi.org/10.1093/trstmh/traa020
893 Monira, S., Nakamura, S., Gotoh, K., Izutsu, K., Watanabe, H., Alam, N., Endtz, H., Cravioto, A., Ali,
894 Sk., Nakaya, T., Horii, T., Iida, T., Alam, M., 2011. Gut Microbiota of Healthy and
895 Malnourished Children in Bangladesh. Frontiers in Microbiology 2.
896 Olm, M.R., Brown, C.T., Brooks, B., Banfield, J.F., 2017. dRep: a tool for fast and accurate genomic
897 comparisons that enables improved genome recovery from metagenomes through de-
898 replication. ISME J 11, 2864–2868. https://doi.org/10.1038/ismej.2017.126

899 Page, A.J., Taylor, B., Delaney, A.J., Soares, J., Seemann, T., Keane, J.A., Harris, S.R., 2016. SNP-sites:
900 rapid efficient extraction of SNPs from multi-FASTA alignments. Microb Genom 2, e000056.
901 https://doi.org/10.1099/mgen.0.000056
902 Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W., 2015. CheckM: assessing the
903 quality of microbial genomes recovered from isolates, single cells, and metagenomes.
904 Genome Res. 25, 1043–1055. https://doi.org/10.1101/gr.186072.114
905 Pérez-Burillo, S., Hinojosa-Nogueira, D., Navajas-Porras, B., Blasco, T., Balzerani, F., Lerma-Aguilera,
906 A., León, D., Pastoriza, S., Apaolaza, I., Planes, F.J., Francino, M.P., Rufián-Henares, J.Á., 2021.
907 Effect of Freezing on Gut Microbiota Composition and Functionality for In Vitro
908 Fermentation Experiments. Nutrients 13, 2207. https://doi.org/10.3390/nu13072207
909 Pop, M., Walker, A.W., Paulson, J., Lindsay, B., Antonio, M., Hossain, M.A., Oundo, J., Tamboura, B.,
910 Mai, V., Astrovskaya, I., Bravo, H.C., Rance, R., Stares, M., Levine, M.M., Panchalingam, S.,
911 Kotloff, K., Ikumapayi, U.N., Ebruke, C., Adeyemi, M., Ahmed, D., Ahmed, F., Alam, M.T.,
912 Amin, R., Siddiqui, S., Ochieng, J.B., Ouma, E., Juma, J., Mailu, E., Omore, R., Morris, J.G.,
913 Breiman, R.F., Saha, D., Parkhill, J., Nataro, J.P., Stine, O.C., 2014. Diarrhea in young children
914 from low-income countries leads to large-scale alterations in intestinal microbiota
915 composition. Genome Biology 15, R76. https://doi.org/10.1186/gb-2014-15-6-r76
916 Porras, A.M., Brito, I.L., 2019. The internationalization of human microbiome research. Current
917 Opinion in Microbiology 50, 50–55. https://doi.org/10.1016/j.mib.2019.09.012
918 Poulsen, C.S., Kaas, R.S., Aarestrup, F.M., Pamp, S.J., 2021. Standard Sample Storage Conditions Have
919 an Impact on Inferred Microbiome Composition and Antimicrobial Resistance Patterns.
920 Microbiology Spectrum 9, e01387-21. https://doi.org/10.1128/Spectrum.01387-21
921 Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013.
922 The SILVA ribosomal RNA gene database project: improved data processing and web-based
923 tools. Nucleic Acids Research 41, D590–D596. https://doi.org/10.1093/nar/gks1219
924 Robertson, R.C., 2020. The Gut Microbiome in Child Malnutrition, in: Michaelsen, K.F., Neufeld, L.M.,
925 Prentice, A.M. (Eds.), Nestlé Nutrition Institute Workshop Series. S. Karger AG, pp. 133–144.
926 https://doi.org/10.1159/000503352
927 Robertson, R.C., Manges, A.R., Finlay, B.B., Prendergast, A.J., 2019. The Human Microbiome and
928 Child Growth – First 1000 Days and Beyond. Trends in Microbiology 27, 131–147.
929 https://doi.org/10.1016/j.tim.2018.09.008
930 Roesch, L.F.W., Casella, G., Simell, O., Krischer, J., Wasserfall, C.H., Schatz, D., Atkinson, M.A., Neu, J.,
931 Triplett, E.W., 2009. Influence of Fecal Sample Storage on Bacterial Community Diversity.
932 Open Microbiol J 3, 40–46. https://doi.org/10.2174/1874285800903010040

933 Sabnis, A., Carrasco, R., Liu, S.X.L., Yan, X., Managlia, E., Chou, P.M., Tan, X.-D., De Plaen, I.G., 2015.
934 Intestinal Vascular Endothelial Growth Factor Is Decreased in Necrotizing Enterocolitis.
935 Neonatology 107, 191–198. https://doi.org/10.1159/000368879
936 Seemann, T., 2023. ABRicate: Mass screening of contigs for antimicrobial and virulence genes.
937 Siafakas, C.G., Anatolitou, F., Fusunyan, R.D., Walker, W.A., Sanderson, I.R., 1999. Vascular
938 endothelial growth factor (VEGF) is present in human breast milk and its receptor is present
939 on intestinal epithelial cells. Pediatr Res 45, 652–657. https://doi.org/10.1203/00006450-
940 199905010-00007
941 Surono, I.S., Widiyanti, D., Kusumo, P.D., Venema, K., 2021. Gut microbiota profile of Indonesian
942 stunted children and children with normal nutritional status. PLOS ONE 16, e0245399.
943 https://doi.org/10.1371/journal.pone.0245399
944 Taft, D.H., Lewis, Z.T., Nguyen, N., Ho, S., Masarweh, C., Dunne-Castagna, V., Tancredi, D.J., Huda,
945 M.N., Stephensen, C.B., Hinde, K., von Mutius, E., Kirjavainen, P.V., Dalphin, J.-C., Lauener,
946 R., Riedler, J., Smilowitz, J.T., German, J.B., Morrow, A.L., Mills, D.A., 2022. Bifidobacterium
947 Species Colonization in Infancy: A Global Cross-Sectional Comparison by Population History
948 of Breastfeeding. Nutrients 14, 1423. https://doi.org/10.3390/nu14071423
949 Tonkin-Hill, G., MacAlasdair, N., Ruis, C., Weimann, A., Horesh, G., Lees, J.A., Gladstone, R.A., Lo, S.,
950 Beaudoin, C., Floto, R.A., Frost, S.D.W., Corander, J., Bentley, S.D., Parkhill, J., 2020.
951 Producing polished prokaryotic pangenomes with the Panaroo pipeline. Genome Biology 21,
952 180. https://doi.org/10.1186/s13059-020-02090-4
953 van Aartsen, J.J., Moore, C.E., Parry, C.M., Turner, P., Phot, N., Mao, S., Suy, K., Davies, T., Giess, A.,
954 Sheppard, A.E., Peto, T.E.A., Day, N.P.J., Crook, D.W., Walker, A.S., Stoesser, N., 2019.
955 Epidemiology of paediatric gastrointestinal colonisation by extended spectrum
956 cephalosporin-resistant Escherichia coli and Klebsiella pneumoniae isolates in north-west
957 Cambodia. BMC Microbiol 19, 59. https://doi.org/10.1186/s12866-019-1431-9
958 Vatanen, T., Ang, Q.Y., Siegwald, L., Sarker, S.A., Le Roy, C.I., Duboux, S., Delannoy-Bruno, O., Ngom-
959 Bru, C., Boulangé, C.L., Stražar, M., Avila-Pacheco, J., Deik, A., Pierce, K., Bullock, K., Dennis,
960 C., Sultana, S., Sayed, S., Rahman, M., Ahmed, T., Modesto, M., Mattarelli, P., Clish, C.B.,
961 Vlamakis, H., Plichta, D.R., Sakwinska, O., Xavier, R.J., 2022. A distinct clade of
962 Bifidobacterium longum in the gut of Bangladeshi children thrives during weaning. Cell 185,
963 4280-4297.e12. https://doi.org/10.1016/j.cell.2022.10.011
964 Vuorela, P., Andersson, S., Carpén, O., Ylikorkala, O., Halmesmäki, E., 2000. Unbound vascular
965 endothelial growth factor and its receptors in breast, human milk, and newborn intestine.

The American Journal of Clinical Nutrition 72, 1196–1201. https://doi.org/10.1093/ajcn/72.5.1196

Zanardo, V., Nicolussi, S., Cavallin, S., Trevisanuto, D., Barbato, A., Faggian, D., Favaro, F., Plebani, M., 2005. Effect of Maternal Smoking on Breast Milk Interleukin-1α, β-Endorphin, and Leptin Concentrations. Environ Health Perspect 113, 1410–1413. https://doi.org/10.1289/ehp.7702

Zani, G., Biavati, B., Crociani, F., Matteuzzi, D., 1974. Bifidobacteria from the Faeces of Piglets. Journal of Applied Bacteriology 37, 537–547. https://doi.org/10.1111/j.1365-2672.1974.tb00479.x

Supplementary Files

Supplementary file 1. Genus differences between DNAShield and Frozen storage conditions.

Supplementary File 2. Metadata and Genus Sequence Data

Supplementary file 3. Genus data Maaslin2 analysis results.

Supplementary file 4. Genus prevalence in samples.

Supplementary file 5. Combined sample metadata and species sequencing data for all samples.

Supplementary file 6. Species data Maaslin2 analysis results.

Supplementary file 7. Species prevalence in samples.

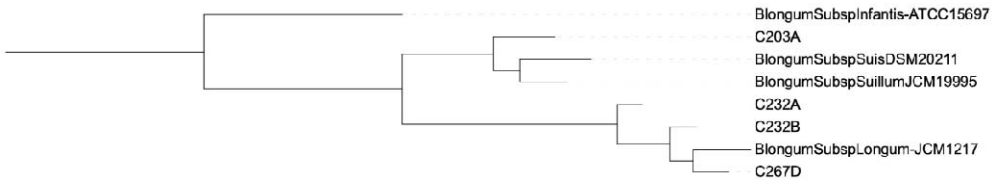
Supplementary File 8. Pairwise Average Nucleotide Identity (ANI)

Supplementary file 9. Metadata and Faecal Cytokine Concentrations.

Supplementary file 10. MaAsLin2 Cytokine Results

991 **Supplementary Figures:**

992

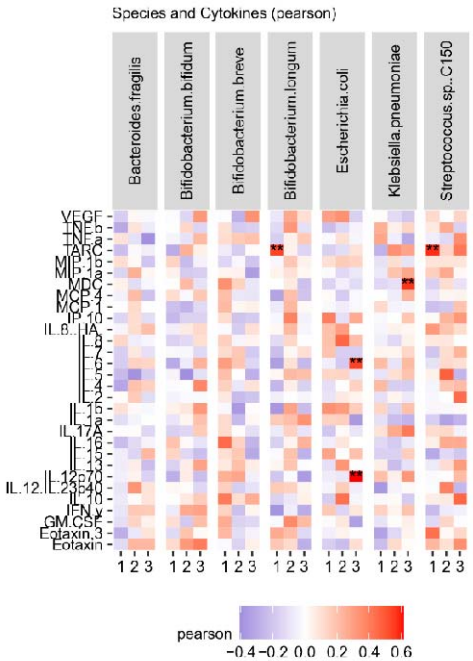


993

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

996

997



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