The, Hao Chung; Florez de Sessions, Paola; Jie, Song; Pham Thanh, Duy; Thompson, Corinne N; Nguyen Ngoc Minh, Chau; Chu, Collins Wenhan; Tran, Tuan-Anh; Thomson, Nicholas R; Thwaites, Guy E; +3 more... Rabaa, Maia A; Hibberd, Martin; Baker, Stephen; (2017) Assessing gut microbiota perturbations during the early phase of infectious diarrhea in Vietnamese children. Gut microbes, 9 (1). pp. 38-54. ISSN 1949-0976 DOI: https://doi.org/10.1080/19490976.2017.1361093

Downloaded from: http://researchonline.lshtm.ac.uk/id/eprint/4189881/

DOI: https://doi.org/10.1080/19490976.2017.1361093

Usage Guidelines:

Please refer to usage guidelines at https://researchonline.lshtm.ac.uk/policies.html or alternatively contact researchonline@lshtm.ac.uk.

Available under license: http://creativecommons.org/licenses/by-nc-nd/2.5/
Assessing gut microbiota perturbations during the early phase of infectious diarrhea in Vietnamese children

Hao Chung The, Paola Florez de Sessions, Song Jie, Duy Pham Thanh, Corinne N. Thompson, Chau Nguyen Ngoc Minh, Collins Wenhan Chu, Tuan-Anh Tran, Nicholas R. Thomson, Guy E. Thwaites, Maia A. Rabaa, Martin Hibberd & Stephen Baker

To cite this article: Hao Chung The, Paola Florez de Sessions, Song Jie, Duy Pham Thanh, Corinne N. Thompson, Chau Nguyen Ngoc Minh, Collins Wenhan Chu, Tuan-Anh Tran, Nicholas R. Thomson, Guy E. Thwaites, Maia A. Rabaa, Martin Hibberd & Stephen Baker (2017): Assessing gut microbiota perturbations during the early phase of infectious diarrhea in Vietnamese children, Gut Microbes, DOI: 10.1080/19490976.2017.1361093

To link to this article: https://doi.org/10.1080/19490976.2017.1361093

© 2017 The Author(s). Published with license by Taylor & Francis® Hao Chung The, Paola Florez de Sessions, Song Jie, Duy Pham Thanh, Corinne N. Thompson, Chau Nguyen Ngoc Minh, Collins Wenhan Chu, Tuan-Anh Tran, Nicholas R. Thomson, Guy E. Thwaites, Maia A. Rabaa, Martin Hibberd, and Stephen Baker

Accepted author version posted online: 02 Aug 2017. Published online: 02 Aug 2017.

Article views: 317
Assessing gut microbiota perturbations during the early phase of infectious diarrhea in Vietnamese children

Hao Chung The, Paola Florez de Sessions, Song Jie, Duy Pham Thanh, Corinne N. Thompson, Chau Nguyen Ngoc Minh, Collins Wenhan Chu, Tuan-Anh Tran, Nicholas R. Thomson, Guy E. Thwaites, Maia A. Rabaa, Martin Hibberd, and Stephen Baker.

ABSTRACT
Diarrheal diseases remain the second most common cause of mortality in young children in developing countries. Efforts have been made to explore the impact of diarrhea on bacterial communities in the human gut, but a thorough understanding has been impeded by inadequate resolution in bacterial identification and the examination of only few etiological agents. Here, by profiling an extended region of the 16S rRNA gene in the fecal microbiome, we aimed to elucidate the nature of gut microbiome perturbations during the early phase of infectious diarrhea caused by various etiological agents in Vietnamese children. Fecal samples from 145 diarrheal cases with a confirmed infectious etiology before antimicrobial therapy and 54 control subjects were analyzed. We found that the diarrheal fecal microbiota could be robustly categorized into 4 microbial configurations that either generally resembled or were highly divergent from a healthy state. Factors such as age, nutritional status, breastfeeding, and the etiology of the infection were significantly associated with these microbial community structures. We observed a consistent elevation of Fusobacterium mortiferum, Escherichia, and oral microorganisms in all diarrheal fecal microbiome configurations, proposing similar mechanistic interactions, even in the absence of global dysbiosis. We additionally found that Bifidobacterium pseudocatenulatum was significantly depleted during dysenteric diarrhea regardless of the etiological agent, suggesting that further investigations into the use of this species as a dysentery-orientated probiotic therapy are warranted. Our findings contribute to the understanding of the complex influence of infectious diarrhea on gut microbiome and identify new opportunities for therapeutic interventions.

Introduction
Diarrheal diseases result in approximately 1.7 billion new infections and 0.75 million deaths in children aged under 5 y annually, making it the second most common cause of mortality in young children in developing countries. Repeated diarrheal episodes cumulatively increase the risk of malnutrition and stunting, which are associated with cognitive impairment and development of cardiovascular diseases and glucose intolerance in adulthood. Such syndromes exhaust societal resources, especially in impoverished regions; therefore there is a substantial demand for efficacious treatments and prophylaxes.

The human gastrointestinal tract is populated with an immensely rich and diverse microbial communities, with the large intestine harboring the greatest density of bacteria. Studies regarding the gut microbiota have highlighted the impact that microbial communities exert on human health, specifically in relation to nutrition, metabolic diseases, and cancer. Various techniques, including metagenomics, have been used to investigate microbial disturbances.
in persistent *Clostridium difficile* infections\textsuperscript{11} and inflammatory bowel disease,\textsuperscript{12} but such dysbiosis remains insufficiently characterized for infectious diarrhea in high incidence settings. The highly dynamic succession of microbial colonization in young children confounds analysis in this target group.\textsuperscript{13,14} Nevertheless, several 16S rRNA gene profiling studies have shown consistent patterns in the initial gut microbiota response following acute diarrheal episodes, in which a transition toward Proteobacteria and *Streptococcus* is observed in an increasingly oxygenated environment.\textsuperscript{15–17} This shift is additionally coupled with a reduction in specific Firmicutes and Bacteroidetes colonizers; their relative abundance is restored during post-diarrheal recovery state.\textsuperscript{15,17–19} These previous studies have either focused on diarrhea associated with *Vibrio cholerae*,\textsuperscript{15,18} or did not offer sufficiently detailed granularity for understanding microbiome alternation.\textsuperscript{16,19} Here, we aimed to characterize modifications of the gut microbiota in the early phase of infectious diarrhea in Vietnamese children by examining their fecal bacterial 16S rRNA composition.

**Results**

**16S rRNA gene sequencing of diarrheal fecal samples**

We subjected a collection of 200 fecal samples (55 samples from non-diarrheal controls and 145 samples from diarrheal patients) from Vietnamese children, to DNA extraction and 16S rRNA gene sequencing (Table 1 and Table S1). 15/54 controls showed presence of at least one pathogen after detection, including Norovirus, Rotavirus, *Salmonella*, and *Campylobacter*. However, the isolation of these organisms in asymptomatic carriers is frequently observed in endemic setting.\textsuperscript{20} One hundred and ninety nine of the fecal samples produced 16S rRNA sequences (median library size = 2,880,968 paired-end reads). Subsampling and quality filtering generated sub-libraries with a median depth of 756,829 paired-end reads, which served as the inputs for EMIRGE assemblies of the V3-V6 region. A mean of 74% of the reads (range: 37% to 92.5%) were successfully mapped and retrieved. A total of 131,702 assembled and clustered sequences, representing a pool of individual sample OTUs (operational taxonomic units), were produced from all samples. Subsequent filtering, OTU reconstruction, and chimeric sequence removal resulted in 7,479 OTUs clustering at 97% similarity in the 199 fecal samples.

**Microbial structures within the fecal samples**

We examined the microbial composition of the fecal specimens from controls and diarrheal cases using the relative abundance of 205 agglomerated unique genera and their phylogenetic relationships. The average relative abundance of the various OTUs in the fecal specimens, up to genus level, is shown in Fig. S1. De novo clustering of the weighted Unifrac dissimilarity matrix, representing the pairwise diversity between samples, segregated the samples into 4 community state types (CSTs). The optimal number of CST groups was determined by gap statistics (Fig. S2) and supported by a prediction strength of > 75%. A principal coordinate analysis (PCoA) using the weighted Unifrac matrix provided further support for the structuring of the 4 different microbial CSTs (Fig. S3). Constrained analysis on the principal coordinates of 195 samples with their associated metadata indicated that age, weight-for-age Z (WAZ) score, and disease status (diarrhea/asymptomatic) best explained the variation between CSTs (Fig. 1). A random forest algorithm was used to reclassify each sample to one of the 4 CSTs based on the relative abundance of 205 genera, yielding a predicted error rate of 11.06%. Twenty-eight genera with the highest calculated variable importance measurement (> 0.002) were selected as the most significant taxa segregating the 4 CSTs. Random forest classification was then repeated with these taxa only to re-assess the most predictive genera for CST clustering, yielding a similar misclassification error rate of

---

**Table 1. Summary of the etiologies associated with 142 diarrheal samples contributing to this study.**

<table>
<thead>
<tr>
<th>Etiological agent</th>
<th>Number</th>
<th>Proportion (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Viral</td>
<td>39</td>
<td>27.5</td>
</tr>
<tr>
<td>Norovirus (NoV)</td>
<td>19</td>
<td>13.4</td>
</tr>
<tr>
<td>Rotavirus (RoV)</td>
<td>20</td>
<td>14.1</td>
</tr>
<tr>
<td>Bacterial</td>
<td>80</td>
<td>56.3</td>
</tr>
<tr>
<td><em>Campylobacter</em></td>
<td>14</td>
<td>9.9</td>
</tr>
<tr>
<td><em>Salmonella</em></td>
<td>15</td>
<td>10.6</td>
</tr>
<tr>
<td><em>Shigella</em></td>
<td>49</td>
<td>34.5</td>
</tr>
<tr>
<td><em>Plesiomonas</em></td>
<td>4</td>
<td>2.8</td>
</tr>
<tr>
<td>Mixed</td>
<td>23</td>
<td>16.2</td>
</tr>
<tr>
<td><em>Campylobacter</em> + NoV/RoV</td>
<td>7</td>
<td>4.9</td>
</tr>
<tr>
<td><em>Plesiomonas</em> + NoV/RoV</td>
<td>1</td>
<td>0.7</td>
</tr>
<tr>
<td><em>Salmonella</em> + NoV/RoV</td>
<td>4</td>
<td>2.8</td>
</tr>
<tr>
<td><em>Shigella</em> + NoV/RoV</td>
<td>11</td>
<td>7.7</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>142</td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>
11.56% (Fig. S4A and B). The following taxa predominated in each of the CSTs (with species included based on their abundances): CST1, Bifidobacterium-dominant (B. longum, B. breve, B. pseudocatenulatum, B. bifidum); CST2, Bacteroides-dominant (B. fragilis, B. vulgatus, B. uniformis, B. caccae); CST3, Streptococcus-dominant (S. gallolyticus and S. salivarius); and CST4, Escherichia-dominant (Fig. 2). A multivariate homogeneity of group dispersions test showed that CST3 possessed the highest measurement. This high variability was explained by the fact that several samples from this CST were rich in Megamonas, as opposed to Streptococcus. For diarrheal cases, the median reported duration of diarrhea onset was 2 d (IQR: 1–3 days). Thus, we classified our findings as representing the early phase of infectious diarrhea since the gut microbiota composition differs as diarrhea progresses to later stages.17,18

The majority of the asymptomatic samples were within CST1 (n = 27) and CST2 (n = 22), which were generally enriched for taxa known to correspond with a healthy gut microbiota in young children.13,21 CST2 harbored a greater degree of bacterial diversity than CST1 and encompassed various genera that delineate an increasingly mature gut microbiota, including Faecalibacterium, Prevotella, Clostridium,
Lachnospiraceae, and *Phascolarctobacterium*.* The fecal samples from the diarrheal cases were distributed across the 4 CSTs, with CST3 (n = 52), and CST4 (n = 29) comprised almost exclusively of diarrheal cases. This structure corresponded with the dominance of *Streptococcus* and *Escherichia* species in gut microbiota of diarrheal children, as reported previously. The estimated Shannon diversity index was elevated in both controls and the cases in CST2 in comparison to the other CSTs (p < 0.05, ANOVA-Tukey's test) (Fig. 3). We additionally observed that bacterial diversity was lower in the diarrheal cases relative to the controls within CST1 and CST2, however this difference between the groups was not statistically significant (p > 0.05, ANOVA-Tukey's test).

**Factors affecting microbial structure in the feces of children with diarrhea**

The CST grouping of the samples from diarrheal children (n = 136), and their respective demographic and clinical metadata, were input into a multinomial logistic regression model to identify factors explaining the grouping of the 4 CSTs (Table 2). CST2 was selected as a reference as it represented a mature gut microbiota configuration. We found that age, WAZ score, infection type (bacterial, viral, or mixed infection (Table 1), and feeding practices remained significantly associated with CST classification (p < 0.05; likelihood ratio test). Furthermore, younger age was significantly associated with CST1 (median age = 16 months; p = 0.015; 2-tailed Z test) and CST3 (median age = 15 months; p = 0.042; 2-tailed Z test) (Table S2), while the fecal samples in CST2 and CST4 were generally associated with older children (median age = 22 and 23 months, respectively). This age differential may be explained by the maturation of the gastrointestinal microbiota in children as the *Bifidobacterium*-rich stage (CST1) precedes the *Bacteroides*-Firmicutes stage (CST2). Additionally, we found that a lower WAZ score was associated with CST4 (median WAZ score = −1.1; p = 0.008; 2-tailed Z test), while CST3 contained a significantly greater proportion of fecal samples from children infected with bacterial diarrheal pathogens (73%; p = 0.001; 2-tailed Z test; Table S2). Lastly, children that were exclusively breastfed were significantly associated with CST1 (p = 0.015; 2-tailed Z test). However, there was no significant difference in feeding patterns according to age (p = 0.422, Kruskal-Wallis rank sum test), suggesting that the association of breastfeeding with CST1 was not a product of clustering with younger age group.

**Figure 3.** Shannon diversity index for 195 fecal samples, classified by diarrheal status and CST membership. Boxplots showing the Shannon diversity index, the upper whisker extends from the 75th percentile to the highest value within the 1.5 *IQR* of the hinge, the lower whisker extends from the 25th percentile to the lowest value within 1.5 *IQR* of the hinge. Data points beyond the end of the whiskers are outliers. The asterisk indicates statistical significant in the pairwise comparison between CST2 and the other CSTs (ANOVA-Tukey's test, p < 0.05).
Changes in taxonomic abundance in the feces of children with diarrhea

Streptococcus and Escherichia species were highly abundant in the diarrheal gut communities associated with CST3 and CST4, respectively. However, 64/145 (44%) of the fecal bacterial community structures from diarrheal patients were more similar to those from the control groups (CSTs 1 and 2). We used DESeq2 to identify the OTUs that were differentially abundant in the diarrheal samples of all 4 CSTs, taking into account the overarching discrepancy in composition due to different CST membership, aiming to minimize effects associated with confounding factors such as age, WAZ scores, and infection type. Seventeen OTUs were found to be at least 4-fold higher in abundance in diarrheal samples than control samples ($p < 0.05$); the majority of these were Escherichia species or organisms associated with the oral cavity, including Fusobacterium, Gemella, and Actinomyces (Fig. 4A). The most significant difference in a bacterial species in diarrheal fecal samples was associated with Fusobacterium mortiferum, which was 210 fold more abundant in diarrheal fecal samples than control samples ($p < 0.0001$). Furthermore, Sutterella, Megamonas, and Enterococcus species were also in higher abundance in diarrheal fecal samples in comparison to control samples. This analysis also identified 19 OTUs that were significantly depleted in diarrheal samples ($p < 0.05$). The majority of these OTUs (13/19) were gut colonizers belonging to the orders Clostridiales and Erysipelotrichales, including Subdoligranulum, Roseburia, Eubacterium, Coprococcus, Catenibacterium, and Ruminococcaceae. Among these, the most depleted taxon was the short chain fatty acid (SCFA) producer Blautia hansenii, which was significantly reduced in samples from those with diarrhea, regardless of the overall community group of the sample ($p < 0.0001$) (Fig. S5). B. bifidum, B. uniformis, and L. reuteri were also significantly associated with a healthy microbial configuration, albeit with different proportions in the various CSTs, likely reflecting age-specific development of the gut microbiota (Fig. S5). In addition, we performed the same analysis on a subset of samples belonging to CSTs 1 and 2, and this produced comparable results to the aforementioned.

We next constructed a correlation network that encompassed representative fecal taxa from both control and diarrheal samples. The resulting correlation network formed a tightly connecting interaction between members of the Clostridiales, Erysipelotrichales, and Bacteroidales, which indicated a positive interaction between these colonizers of a healthy gut (Fig. 5). A further positive interacting cluster was composed of genera that are frequently associated with the human oral cavity, including Fusobacterium nucleatum, Parvimonas micra, Peptostreptococcus stomatis,
Solobacterium moorei, Gemmella haemolysans, Actinomyces odontolyticus, Lachnoanaerobaculum orale, Abiotrophia defectiva, Atopobium parvulum, Rothia mucilaginosa, and Streptococcus salivarius (Fig. 5). Conversely, Rothia was negatively correlated with Bacteroides vulgatus and Faecalibacterium prausnitzii, which are both indicators of a healthy and mature gastrointestinal microbiota.8,13 We additionally found that Fusobacterium mortiferum, the significantly enriched taxon in fecal samples from diarrheal children, exhibited a negative association with a key component of the gut Clostridiales network (Clostridium innocuum). This suggests that the colonization of F. mortiferum may inhibit the proliferation of related dependent taxa such as Blautia, Ruminococcus, and several Bacteroides. The negative correlation observed between Bifidobacterium longum subsp. infantis and Bifidobacterium breve likely reflects a competitive ecology as the gut microbiota matures.22

**Figure 4.** Bacterial taxa showing significantly different abundance among the examined classes. OTUs were identified to be of significantly differential abundance between groups in examination, as detected and filtered by DESeq2. In short, only OTUs with adjusted p values < 0.05, estimated fold change > 4 or < 1/4, and estimated base mean > 30 were considered significantly differentially abundant and included in the plot. (A) Ratio of the log2 fold change of OTUs that differ between diarrheal and control stools, accounting for the different general microbial configurations in 4 CSTs (N = 199). (B) Ratio of the log2 fold change of OTUs that differ between bacterial and viral diarrheal infections, accounting for the different general microbial configurations in 4 CSTs (N = 119). (C) Relative abundances of OTUs that differ between dysenteric and non-dysenteric stools, accounting for different types of infection (N = 142).

**Taxonomic changes associated with bacterial induced diarrhea and dysentery**

To explore the differential composition of the gut microbiota in response to the different etiological agents (viral vs. bacterial), we examined a subset of 119 diarrheal fecal samples (viral infections = 39; bacterial infections = 80) to eliminate potential misclassification induced by 23 cases of mixed infections. Analysis using DESeq2 identified a group of 19 OTUs that were significantly more abundant in bacterial infections across the 4 CSTs. Among these, 8 taxa were more commonly associated with human oral microbiota (Fig. 4B); Actinomyces and Rothia were the most prevalent (assessed
by the DESeq2 calculated base mean value). Other organisms with high relative proportion in the fecal samples from children with bacterial induced diarrhea included \textit{C. ramosum} and \textit{B. vulgatus}. Conversely, 4 OTUs were significantly elevated in the viral diarrheal infections; these included the \textit{Bifidobacterium} species, \textit{B. breve} and \textit{B. pseudocatenulatum} (Fig. 4B).

We lastly categorized 142 diarrheal fecal samples based on the diagnosis of dysentery (the visible presence of mucoid and/or blood in the stool; \( n = 60 \)), accounting for the differential microbiota changes due to types of infectious agents (bacterial, viral, or mixed infection). We identified 2 OTUs that were significantly associated with fecal samples from non-dysenteric diarrhea (DESeq2, \( p < 0.05 \)) (Fig. 4C). \textit{B. pseudocatenulatum} was almost \( 2^5 \) fold less abundant in dysenteric feces in comparison to non-dysenteric diarrheal feces; this was consistent across all of the corresponding etiological agents (Fig. 4C). For bacterial infections, \textit{L. reuteri} was uniformly more depleted in the examined dysenteric fecal samples, but its variability was higher in dysenteric stools in those with a viral or mixed infection. When we accounted only for the stool with visible blood (\( n = 10 \)), the analysis revealed that \textit{S. gallolyticus} was the most depleted taxon in dysentery compared with watery diarrhea (DESeq2 Wald test without effect of infection types, \( p < 0.05 \)). Other taxa of decreased abundance included \textit{Escherichia}, \textit{B. pseudocatenulatum}, \textit{Enterococcus} and several oral microbiome species. However, these interpretations are to be confirmed in larger studies.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{correlation_network}
\caption{Correlation network of the healthy and diarrheal microbiome. Figure shows the correlation network of the 92 most represented OTUs sampled from 199 stool isolates, defined as OTUs with occurrence in at least 10 samples, constructed using the SparCC wrapper from package ‘SpiecEasi’. Only correlations with calculated \( p \) value \( < 0.05 \) and absolute magnitude \( \geq 0.25 \) were shown in the network. Positive and negative interactions were denoted as a red and blue solid line respectively, with line weight proportional to correlation strength. The OTUs (nodes) were colored based on taxonomic family (see legend), with sizes proportional to their relative abundances. The light green shaded area covers OTUs identified as members of normal human oral microbiota (through comparison with the Human Oral Microbiome Database).}
\end{figure}
Discussion

Here we assessed the impact of diarrheal infection on the composition of the gut microbiota in Vietnamese children. Our results predict that the gut microbiota in young children exhibits differential responses during the early phase of diarrhea, which can be grouped into microbial community structures that either closely resemble or are highly divergent from those of healthy children. Approximately 44% of the diarrheal cases in this study retained a Bifidobacterium- or Bacteroides-rich structure. This association largely resembles the well-studied age-dependent normal gut microbiota in children,14 outlining a lack of global dysbiosis in these children. Among the various perturbed microbial states, we found that age, nutritional status, breastfeeding, and diarrheal etiology contributed to the composition of the bacterial communities during the early phase of diarrhea. However, since the WAZ score was collected after the onset of diarrhea, weight could be subjected to temporary fluctuation and its surrogate for nutritional status should be interpreted with caution. A recent study on diarrheal microbiome from U.S patients reported that diarrheal microbial structures could be grouped into 4 major clusters irrespective to the etiologies, of which 2 are almost exclusively restricted to diarrheal patients and associated with higher prevalence of Escherichia.23 In our study, Streptococcus was associated with younger age and bacterial infection, while Escherichia was associated with older children and a poor nutritional status. Indeed, analysis of the microbiota in diarrheal samples from young children (<2 y old) predominantly infected with pathogenic E. coli revealed that Streptococcus (S. gallolyticus and S. salivarius) was the most abundant organisms during the early phase of diarrhea.24 These 2 species were also dominant in our study. In contrast, Escherichia were overrepresented in the fecal samples from children (2–3 y old) infected with cholera in Bangladesh.15 Due to limitations of the study, we were unable to assess how these initial differential microbial communities affect diarrheal severity and recovery. Dysentery and antimicrobial treatment may further confound these associations. However, the ratio of Streptococcus/Bifidobacterium showed a significant positive correlation with diarrheal fecal output and duration of hospitalization, suggesting that the higher degree of divergence from the healthy microbiota may predict a more severe clinical presentation.24 Here, for bacterial induced diarrhea cases, we attempted to identify the pathogens’ attributed sequenced reads in the fecal samples. However, only one Salmonella and one Campylobacter infections showed detectable 16S rRNA sequences classified to the corresponding pathogens. Shigella infections were not considered since its sequences are indistinguishable from Escherichia. This concurs with previous finding that by using 16S rRNA profiling on cholera fecal samples, V. cholerae is only detectable during the first day of symptom onset, and its presence is significantly reduced even on the second day of diarrhea.17

We additionally observed that obligate anaerobes belonging to the orders Clostridiales and Erysipelotrichales were consistently depleted in the diarrheal cases in comparison to the controls. This group included multiple SCFA producers including Blautia, Subdoligranulum, Roseburia, and Eubacterium. This finding may explain the transient depletion of SCFA following diarrhea, resulting in poor water and electrolyte absorption and a shortage of metabolic energy for enterocytes, which may lead to dehydration and further complications.25 These data are in accordance with previous studies, which have shown that Clostridiales become gradually enriched during the diarrheal recovery period.15,17,18 This loss of obligate anaerobes is coupled with the proliferation of facultative anaerobes, such as the Streptococcus and Escherichia species.16,17 Unpredictably, we observed that a substantial component of the microbiota from the human oral cavity was consistently represented in the diarrheal fecal samples. We propose that the transiently oxygenated environment and potential decrease in bacterial competition in the diarrheal gut enhances the colonization potential of these species.17 This may explain why S. salivarius, a dominant taxon of the oral cavity of young children, is capable of establishing an exceptionally high rate of colonization in suckling infants. Oral commensal organisms, predominantly S. salivarius and F. nucleatum, may serve as nucleators for the adherence of other bacteria, eventually forming polymicrobial biofilms that are capable of surviving mechanical and chemical stresses during their passage through the large intestine.26,27 Alternatively, the diarrheal-associated obligate anaerobe, F. mortiferum, which was found not to be associated with oral organisms, exhibited antagonism with the Clostridiales. This finding is in agreement with the spike of
Fusobacterium observed in early phase diarrhea.\textsuperscript{18,19} Unlike the majority of Fusobacterium, F. mortiferum is extremely pleomorphic and bile resistant, with the ability to rapidly accumulate and metabolize a broad spectrum of sugars independent of amino acid fermentation.\textsuperscript{28-30} Furthermore, poultry isolated F. mortiferum has been shown to secrete a bacteriocin that can inhibit the growth of Gram-positive Bacillus species.\textsuperscript{31} Taken together, this evidence implies that F. mortiferum is more resilient to environmental stresses and able to occupy the anaerobic niche of the gut, possibly accelerating the depletion of Clostridiales via bacteriocin-mediated competition.

Dysentery denotes a more severe form of diarrhea, frequently requiring antimicrobial treatment and resulting in longer periods of hospitalization.\textsuperscript{20,32} To date, only Lactobacillus ruminis has been found to be depleted in dysenteric fecal samples.\textsuperscript{16} Here, we found that L. reuteri was significantly diminished in dysenteric diarrhea; we speculate that this discrepancy in species level identification may be induced by variation in environmental and genetic factors of the studied populations. In addition, we report here that a Bifidobacterium species, B. pseudocatenulatum, is substantially depleted in fecal specimens following bacterial infection; this effect is amplified during dysentery. Previous research examining the V1-V2 region of the 16S rRNA failed to efficiently identify Bifidobacterium, thus discounting its contribution.\textsuperscript{16} The oral administration of live B. pseudocatenulatum in induced cirrhotic and diabetic mice has been shown to restore the gut integrity and initiate an anti-inflammatory cytokine profile in the intestine, leading to a diminished systematic inflammatory response.\textsuperscript{33-35} Further, this potential probiotic also induces monocyte-derived macrophages to undergo transition into the anti-inflammatory M2 phase.\textsuperscript{36} It has been shown that the colonization of B. pseudocatenulatum or B. breve, but not other Bifidobacterium, protect mice from lethal Shiga toxin producing E. coli (STEC) infections.\textsuperscript{37} Treatment for dysentery is currently focused on antimicrobial therapy, and the usage of broad spectrum antimicrobials quickly depletes the balanced microbiota derived metabolites such as SCFAs, leading to disruption of the gut immune homeostasis.\textsuperscript{38} Therefore, we propose that B. pseudocatenulatum may be a potential dysentery-oriented probiotic candidate to reduce inflammation associated pathological conditions and accelerate the recovery of the microbiota to a healthy state.

Some limitations should be considered in the context of our investigation. First, the EMIRGE mediated read recovery rate was consistently lower in diarrheal samples. This means that a proportion of uncharacterized taxa were unavailable for examination, and it may impact on the true estimation of the diversity within these samples. Second, an exhaustive search for other diarrheagenic etiologies by molecular methods was not possible during this study; this may underestimate the true incidence of mixed infections. We propose that more comprehensive approaches, such as the Luminex xTAG multi-pathogen panel assay, could be incorporated in further studies to improve diagnostic accuracy.\textsuperscript{39} Third, the non-diarrhea controls included are not ideally representative of the healthy population, given that some of these children sought medical help for nutritional concerns. Notwithstanding these limitations, through sequencing an extended region of the 16S rRNA gene, it was possible to obtain high-confidence taxonomic assignment up to the species level, portraying the highly complex gut microbiota in greater resolution. The inclusion of various diarrheal etiologies and the statistical methods used here further permitted us to reveal the general impact of infectious diarrhea on the gut microbiota. Diarrhea cases with unknown etiologies were excluded from this study, and this may underrepresent the true heterogeneity of diarrheal causes and their differing influences on the microbiota. Furthermore, we cannot dismiss that additional underlying mechanisms such as inflammation induced changes in immune response and oxygen gradient,\textsuperscript{40} aside from the physical manifestation of diarrhea, may contribute to observed changes in the microbiota. Regarding the correlation network analysis, it is apparent that SparCC detects fewer negative (competitive relationships) than positive correlations (mutualism and commensalism). Unlike positive correlations, which are more readily associated with the co-occurrence of OTUs, negative correlations are typically inferred by the mutual exclusions of OTUs. SparCC has been shown to possess increased precision in detecting negative correlations in comparison to other methods, but correlation detection is limited to those with greater power and low noise.\textsuperscript{41}

In conclusion, by studying the microbiota of children with diarrhea associated with a diverse range
of confirmed etiological agents, we have generated a comprehensive insight into the effects of infectious diarrhea on the gut microbiota. Therefore, our study provides a pivotal understanding of the impact of infectious diarrhea on the gut microbiota of children, particularly ones in low-middle income urban settings in Southeast Asia. Future work, through longitudinal study designs and the employment of shotgun metagenomics, should address how the initial differential responses in gut bacterial composition may impact disease progression, such as diarrhea clearance and recovery rate of the depleted microbiota.

Materials and methods

Study design, sample collection and microbiological procedures

Samples in this study originated from a previously described prospective observational study of pediatric diarrhea conducted at 3 major hospitals in Ho Chi Minh City (HCMC), Vietnam: Children’s Hospital 1 (CH1), Children’s Hospital 2 (CH2), and the Hospital for Tropical Diseases (HTD). The Oxford Tropical Research Ethics Committee (OxTREC) and the Scientific and Ethics Committee of CH1, CH2, and HTD provided ethical approval for the primary and this subsequent analysis. The general inclusion criteria for both diarrheal cases and non-diarrheal controls were children aged under 5 years, residing within HCMC and reporting no antimicrobial usage within 3 d before hospital admission. Children admitted to the study sites with diarrhea (defined as 3 or more loose stools or at least one bloody loose stool within a 24 hour period) were included as cases. Children who presented for health checks, nutritional, or gastrointestinal issues but reported no diarrheal or respiratory illnesses within 7 d of admission were recruited as controls. For each enrollee, clinical data regarding the symptoms and duration of diarrhea was obtained from a case report form completed by study clinicians, while demographic, feeding behavior and socioeconomic details were provided through a confidential questionnaire. Weight-for-age Z (WAZ) score was used to evaluate the nutritional status of all enrolled children based on WHO standards.

A fecal sample was collected from each participant, both cases and controls, before any prescribed antimicrobial treatment. Both case and control fecal samples were subjected to standard microbiological culturing and biochemical testing to identify common diarrheal bacteria, including Shigella, Salmonella, Campylobacter, Plesiomonas, Yersinia, and Aeromonas, as described previously (https://wwwnc.cdc.gov/EID/article/19/6/11-1862-Techapp1.pdf). For the detection of Rotavirus and Norovirus, reverse transcriptase polymerase chain reaction (RT-PCR) was performed on total viral RNA extracted from all specimens. Further detection of Giardia lamblia, Entamoeba histolytica, and Cryptosporidium cysts were performed by microscopy on fresh fecal samples diluted in phosphate buffered saline. Stool samples were stored at −80°C until being subjected to DNA extraction up to 6 months after initial storage.

DNA extraction and sequencing

The prospective study from which the fecal specimens originated successfully recruited 1,419 cases and 609 controls. A subset of 200 fecal samples was selected for this investigation; these included 55 randomly selected controls and 145 cases from multiple confirmed etiologies. For diarrheal cases, random subsamples from each identified etiology were included to represent the epidemiology reported in our previous prospective study, with emphasis on Shigella as it was the major cause of dysentery (Table 1). Total DNA was extracted from these fecal samples using a phenol-chloroform extraction method. Briefly, 200 µl of the fecal sample was suspended in a solution containing 500 µl of DNA extraction buffer, 50 µl NaCl 200mM, 50 µl SDS 10%, 50 µl of phenol:chloroform:isoamyl (25:24:1), and 0.7 g of 0.5-mm-diameter zirconia/silica beads. Cells were lysed by mechanical disruption with a bead beater for one minute and subjected to 3 rounds of phenol:chloroform extraction. DNA was resuspended in TE buffer supplemented with RNase.

All DNA samples were shipped on dry ice to the Genome Institute of Singapore (GIS) for 16S rRNA sequencing using the GIS Efficient Rapid Microbial Sequencing (GERMS) platform. To obtain high resolution taxonomic identification up to the species level, all samples were PCR amplified using a previously optimized primer set (338F-1061R), which produces >700 bp amplicons covering the 4 variable regions (V3-V6) of the 16S rRNA region and can be assembled in silico to retrieve > 92% of the sequences in the Greengenes database. Long PCR fragments were cleaned using 1X AMPURE beads and randomly
fragmented with Covaris (model LE220) shearing to ~200bp. Library preparation was performed using the GeneRead DNA Library I Core Kit (Qiagen, Germany) according to the manufacturer’s instructions, and the prepared DNA was subjected to sequencing on an Illumina HiSeq2500 platform to produce 75 bp paired-end reads.

**EMIRGE assembly of 16S rRNA amplicons and OTU clustering**

EMIRGE (Expectation Maximization Iterative Reconstruction of Genes from the Environment) is a database dependent assembler built on an iterative expectation-maximization (EM) algorithm, which is used to reconstruct full length 16S rRNA small subunit gene sequences (by simultaneously mapping and clustering) and estimate their relative abundance. The SILVA small subunit (SSU) rRNA database version 119 was filtered to remove potential large subunit (LSU) rRNA sequences, and closely related sequences were clustered at 97% identity by USEARCH; this was achieved by using PhylotFlash v2.0 (https://github.com/hrgv/phyloflash). The resulting database was used as the reference template for EMIRGE. To limit the computational effort related to using the complete library without replacement from each sample library using seqtk (https://github.com/lh3/seqtk). Reads were trimmed using Sickle to remove those with quality <30 and length <60. We inputted the trimmed reads from each subsampled library into the EMIRGE output, a set of assembled and clustered sequences, for each sample indicates its representative OTUs with their estimated abundances. Assembled sequences with sample-wise normalized relative abundances less than 0.01% were removed from further analysis.

A pseudo-count for each sequence per subsampled library was calculated by scaling the number of successfully mapped reads to the EMIRGE estimated relative abundance of each sequence. All filtered sequences and a count table detailing their respective abundance were pooled and imported into the 16S rRNA processing platform mothur v.1.36.0. To minimize the length differences in assembled sequences and to facilitate more accurate OTU clustering, sequences were aligned to a trimmed version of the SILVA reference database to include only the amplified region (338F-1061R). Gap-only columns were filtered from this alignment, and sequences with a maximum of 7 (~1%) ambiguous sites were retained for downstream analyses. Sequences were dereplicated, and ambiguous sites were replaced randomly with one of the 4 nucleotides (ATCG). These unique sequences and their respective counts served as input for the UPARSE clustering algorithm, which clustered all pooled sequences at a 97% similarity threshold.

Chimeric sequences were stringently predicted and removed by UCHIME v6.0, set at both de novo mode and against the ChimeraSlayer informed reference database (e.g. ‘gold’ database). A total of 7,479 OTUs were reconstructed for 199 of the successfully sequenced samples. An alignment consisting of the most abundant representative sequences from each OTU was used to construct a phylogenetic tree using FastTree 2 under default parameters. Taxonomic assignments of OTU representatives up to the genus level were performed using the mothur implemented Ribosomal Database Project (RDP) classifier, with a minimum support threshold of 80%.

**Data analysis**

All analyses were conducted in R using multiple packages, including ‘phyloseq’, ‘cluster’, ‘randomForestSRC’, ‘ggplot2’, ‘nnet’, ‘lmtest’, ‘vegan’, ‘DESeq2’, ‘SpiecEasi’ and other packages. An OTU count table, taxonomy classification table, related clinical and demographic data and the OTU phylogenetic tree were imported and analyzed as a ‘phyloseq’ object, allowing a unified and interactive analysis approach.

**Clustering into community state types (CSTs)**

To gain overall insight into the microbial compositional structure of both controls and diarrheal cases, we applied a previously described clustering approach. To reduce sparsity as well as account for the divergence and the functional similarity shared between members of the same genus, which is frequently reported for human gut microbiota, the OTU count table was agglomerated at the unique genus level for clustering as recommended. Recent benchmark studies have proposed that the weighted Unifrac distance produces desirable accuracy and power in exploring β-diversity. A weighted Unifrac dissimilarity matrix between all samples was calculated using...
the collapsed genera’s relative abundances and their phylogenetic relatedness. The partitioning around meloids (pam) algorithm was applied to this matrix to cluster all samples into distinct CSTs, with the optimal number of CSTs (k = 4) determined using gap statistics, a goodness of clustering measure. Further assessments such as prediction strength, average silhouette width (asw) were used to validate this optimal number of clusters, as recommended previously. Microbial CST grouping has been applied extensively to profile the human vaginal microbiome and has produced consistent findings.

To examine the relationship between the microbiome structures and associated explanatory variables for 195 samples with complete metadata, we applied constrained analysis of principal coordinates (CAP) on the calculated weighted Unifrac dissimilarity matrix and a set of demographic variables (age, sex, WAZ score, feeding pattern, income, rural residence), as well as diarrheal status. This was performed using the ‘capscale’ function in the ‘vegan’ package. Significant variables were identified and included in the final model based on Akaike information criterion (AIC) in a stepwise model selection approach. The α diversity of these 195 samples was estimated by Shannon diversity index. Analysis of variance (ANOVA) with post-hoc Tukey test and Bonferroni correction for multiple comparisons was used to compare control and diarrheal α-diversity among and within each CST. Multinomial logistic regression modeling was applied using the ‘multinom’ function in ‘nnet’ package to evaluate the association of the various aforementioned demographic factors and clinical features (vomiting, dysentery, and infection type) to the CST membership of diarrheal cases, with the Bacteroides rich CST (CST2) serving as the reference group. These predictors were included on the basis of limited missing data and capable of being assessed by clinicians upon patient’s admission. Five samples with WAZ scores >3 or <-3, and one CST1 sample with age exceeding 2 standard deviations were considered as outliers and removed, resulting in 136 diarrheal samples with full metadata being subjected to regression modeling.

Evaluating differential abundance

DESeq2 was used to normalize the 7,479-OTUs table and to detect OTUs that show significantly differential relative abundance between 2 assessed groups per comparison (diarrhea against controls, bacterial infection against viral infection, and dysentery against non-dysentery), as recommended in recent benchmark studies. To identify consistent trends observed across all CSTs (diarrhea against controls, bacterial infection against viral infection) or infection types (dysentery against non-dysentery), we applied the likelihood ratio test (LRT) approach with the reduced models incorporating only either CST or infection type as coefficients. OTUs with adjusted p values < 0.05, estimated fold change >4 or <1/4, and estimated base mean >30 were considered significantly differentially abundant between 2 examined classes. Taxonomic identification to the species level was performed by manually comparing the reconstructed 16S rRNA sequences of OTUs to either the Human Oral Microbial Database (HOMD; http://www.homd.org/) or the Human Microbiome Project (HMP; http://www.hmpdacc.org/resources/blast.php) by BLAST. OTU was assigned a species nomenclature if its sequence is of at least 97% similarity to that in the database. Using these approaches, Shigella sequences are indistinguishable from Escherichia.

Interaction network construction

A correlation network was constructed for all 199 control and diarrheal samples to characterize the potential interactions between most representative 92 OTUs, defined as OTUs detected in at least 10 samples. This filtering step did not substantially affect the representativeness of the data set, with the median sample retention rate of 93% (IQR: 87% - 97%). The network was constructed using the SparCC wrapper in the package ‘SpiecEasi’. The statistical significance for each interaction was assessed by 100 bootstrap iterations, with p values adjusted for multiple comparison correction. To avoid spurious correlations, only
those with adjusted $p$ values no greater than 0.05 and absolute magnitude equal to or above 0.25 were considered as significant correlations and represented in the final plots.

**Abbreviations**

- AIC Akaike Information Criterion
- ANOVA Analysis of Variance
- CST Community State Type
- CAP Constrained analysis of principal coordinates
- EMIRGE Expectation Maximization Iterative Reconstruction of Genes from the Environment
- HCMC Ho Chi Minh City
- IQR Interquartile range
- LRT Likelihood Ratio Test
- OTU Operational Taxa Unit
- SCFA Short Chain Fatty Acid
- WAZ Weight for Age Z score

**Ethical approval**
The Oxford Tropical Research Ethics Committee (OxTREC) and the Scientific and Ethics Committee of Ho Chi Minh city’s Children Hospital 1 (CH1), Children Hospital 2 (CH2), and Hospital for Tropical Diseases (HTD) provided ethical approval for the primary and this subsequent analysis (OxTREC No. 0109).

**Reagents and catalog number**
GeneRead DNA Library I Core Kit (Qiagen, catalog no. 180434)

**Disclosure of potential conflicts of interest**
The authors report no potential conflict of interest.

**Acknowledgements**
We gratefully acknowledge Andreas Sundquist and Matthew Davis, who wrote and modified the C++ script that generated the plot in Supplementary Figure 1. Computational resources were funded by the Li Ka Shing – University of Oxford Global Health Program (LG05, SM27). We express our gratitude to Paul J. McMurdie, Susan P. Holmes and her team for developing the package ‘phylloseq’ and conducting invaluable benchmark studies on analysis of microbiome data.

**Funding**
This work was funded by the Wellcome Trust and the Royal Society (100087/Z/12/Z). DPT is supported by an OAK foundation fellowship.

**ORCID**
Hao Chung [http://orcid.org/0000-0002-4028-4074](http://orcid.org/0000-0002-4028-4074)

**References**


9. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome funded by the Wellcome Trust and the Royal Society (100087/Z/12/Z). DPT is supported by an OAK foundation fellowship.


