

Microbiome Profiling of Enterotoxigenic *Escherichia Coli* (ETEC) Carriers Highlights Signature Differences Between Symptomatic and Asymptomatic Individuals

Ellen Elizabeth Higginson

University of Cambridge <https://orcid.org/0000-0001-9218-1627>

Md Abu Sayeed

International Centre for Diarrhoeal Disease Research Bangladesh

Joana Pereira Dias

University of Cambridge

Vignesh Shetty

University of Cambridge

Mamatha Ballal

Manipal Academy of Higher Education

Sunil Kumar Srivastava

University of Delhi

Firdausi Qadri

International Centre for Diarrhoeal Disease Research Bangladesh

Gordon Dougan

University of Cambridge

Ankur Mutreja (✉ am872@cam.ac.uk)

University of Cambridge <https://orcid.org/0000-0002-1118-8075>

Research

Keywords: ETEC, microbiome, metagenomics

Posted Date: November 9th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1030513/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Escherichia coli (ETEC) are one of the top causes of diarrhoea in children in low- and middle-income countries (LMICs). However, large-scale pathogen burden studies in children have identified ETEC in the guts of symptomatic patients and controls. The factors that influence this balance between carriage and disease are poorly understood, but it is postulated that the gut microbiome may play a role in either resistance or progression to disease. In this study, we investigated the microbiome profiles, using shotgun DNA sequencing, of children and adults from Bangladesh who were asymptotically or symptomatically infected with ETEC.

Results

Symptomatic patients had significantly higher numbers of sequenced reads mapping to both *E. coli* and the two ETEC toxins (LT and ST), suggesting higher bacterial burden. They were also significantly more likely to be co-infected with enteroaggregative *E. coli* (EAEC) and had higher proportions of other Gammaproteobacteria, including *Klebsiella*, *Salmonella*, and *Haemophilus*. Colonisation with ETEC (symptomatic or asymptomatic) was also associated with increased prevalence of antimicrobial resistance (AMR) genes, most notably those of the b-lactamase class. Taxonomic profiles were distinctly different between all groups in both species richness (alpha diversity) and composition (beta diversity), although the direction of these changes was different in adults and children. As seen in previous studies, children with high *E. coli* burdens also had higher proportions of *Streptococcus* spp., while healthy children were more heavily colonised by several *Bifidobacterium* spp.

Conclusions

Our study provides insight into the microbiome changes that occur upon infection with ETEC in an endemic setting, and provides rationale for future studies investigating how the microbiome may protect or predispose individuals to symptomatic infections with gastrointestinal pathogens.

Introduction

Enterotoxigenic *Escherichia coli* (ETEC) are a common cause of watery diarrhoea in low-and-middle income countries (LMICs). While normally described as the major cause of travellers diarrhea in visitors to these regions, ETEC also infect young children in endemic areas, causing an estimated 84.4 million diarrhea episodes per year worldwide (1). Although most people in these regions will generate immunity in childhood, this immune protection is not absolute and adults may periodically endure symptomatic infections. From large-scale field studies, it is also clear that individuals can often carry ETEC without displaying significant clinical symptoms (2). Although it is not known what modulates the balance between symptomatic and asymptomatic infections, factors such as pre-existing immunity and the composition of the microbiome may play a part.

In recent years, there has been increasing interest in understanding how the microbiome impacts human health. However, the majority of this microbial profiling has been performed on healthy adult individuals in developed countries, and may not be generalisable or applicable to people in geographically distinct regions. Exemplar studies have found that the microbiomes of people in the US and Europe are distinct from those in Africa and India (3–5). Some of this can be described by the *Prevotella/Bacteroides* gradient within the Bacteroidetes phylum, whereby *Bacteroides* predominate in Western populations, whilst *Prevotella* dominate elsewhere. Bacteroidetes are also more generally overrepresented in Western microbiomes, with Firmicutes being more common in other populations. This is also true in children, whereby children in Bangladesh were shown to have much lower levels of Bacteroidetes in total, and higher microbial diversity than children in the US (6). Similar patterns were also noted in adult US travellers to Central America and India, with travellers having relatively more Firmicutes and less Bacteroidetes on return to the US (7). These regional variations in microbiome composition are likely to be mediated by a multitude of factors including diet, genetics, lifestyle and the environment (4).

One environmental factor that is expected to have a major influence on gastrointestinal health is the interaction between the microbiome and gastrointestinal pathogens. Although several studies have looked at the impact of intestinal infection on the gut microbiome (8–13), due to differences in patient age, pathogen of interest, control groups and antibiotic therapy, as well as small patient sample numbers, there is no consensus on what aspects of the microbiome may be protective or endow colonization resistance, and which might facilitate progression to disease. In a human ETEC challenge study, researchers identified several species that were correlated with progression to disease or resistance to infection, including several *Bacteroides spp.* which were overrepresented in volunteers who were protected from symptomatic infection (14). However, as this study was in healthy adults in the US, it is not clear how relevant these findings are to populations where ETEC is endemic.

In our study we used metagenomic sequencing analysis to compare the gut microbiomes of adults and children in Bangladesh who were asymptotically or symptomatically infected with ETEC. The whole genome shotgun sequence data were interrogated to determine the burden of *E. coli* species, presence of ETEC toxins, antimicrobial resistance genes (ARGs), other bacterial pathogens and overall microbiome composition.

Methods

Ethics

Samples were collected from participants of the ETEC ETVAX Vaccine Trial (clinical trial NCT02531802) at the International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b).

Sample collection

Stools samples were collected from asymptomatic and symptomatic adults and children who were culture positive for ETEC, as well as uninfected controls (Table 1). Raw stool was stored at -80°C before

total DNA was extracted using the QIAamp DNA Stool Mini Kit. All samples were processed within 4 days of sample collection.

Sequencing

DNA samples that cleared the quality-control check were sequenced in 16-plex on two lanes of a HiSeqV4 machine at the Wellcome Sanger Institute (WSI). The raw data is available on the European Nucleotide Archive under Study Accession PRJEB21793. Over 15 million 125 bp paired-end reads were generated for each sample. The data generated were screened for human reads, which were identified and removed from the raw sequence data by aligning to the *Homo sapiens* GRCh38 reference genome using Bowtie2 v2.2.3 (15).

***E. coli* mapping, virulence factor and ARG detection**

To identify *E. coli* species specific reads in raw sequence data, human-depleted reads were mapped to the widely used O78 reference genome of H10407 (accession no. GCA_000210475.1) using Bowtie2. To identify relevant *E. coli* virulence determinants and ARGs, raw reads were mapped against databases of *E. coli* virulence determinants (Virulence Factor Database (16)) and ARGs (ARGannot3 (REF)) using SRST2 v2 (17). The number of reads mapping to these genes of interest was quantified by calculating the reads per kilobase mapped (RPKM). Pathotypes were defined as follows: Enterotoxigenic *Escherichia coli* (ETEC) were defined as presence of the genes corresponding to heat labile toxin (LT) or heat stable toxin (ST), Enteroaggregative *Escherichia coli* (EAEC) were defined by the presence of *aatA*, and at least one of either the *aggR*, *aaiC* or *aap* genes and Enteropathogenic *Escherichia coli* (EPEC) were defined as *eae* positive and *stx* negative.

Statistical analysis of pathotype detection was calculated using a two-tailed Fishers exact test, and the percent of mapped reads and RPKM values were compared using a Wilcoxon rank-sum test. To compare the prevalence of antibiotic resistance gene detections between groups the Fishers exact test was used with an adjustment for false discovery rate.

Taxonomic classification, abundance estimation and diversity analyses

Classification of human-depleted reads was carried out using Kraken2 v2.0.8-beta (18). Reads were classified against the uhgg_kraken2-db database of prokaryotic genomes (19), which was downloaded from the European Bioinformatics Institute (<https://www.ebi.ac.uk/metagenomics/genomes>). Abundance estimation was run on Kraken2 outputs using Bracken v2.5 (20) at the phylum and genus levels. Diversity analyses were carried out in R using vegan v2.5-6 (21) and agricolae v1.3-1 (22). Before alpha and beta diversity calculations, data was rarefied to the minimum number of reads to maintain all samples in the dataset. Statistical analysis of alpha diversity measures was carried out using ANOVA and Tukey's HSD test.

Metacoder v0.3.3 (23) was used to visualise genus level differences in bacterial diversity between samples with high (> 10%) and low percentages (< 10%) of reads mapping to *E. coli* H10407. Bracken

abundance data was filtered to limit the dataset to bacterial taxa. To discount the contribution of *Escherichia* reads on the overall abundance of taxa, reads classified to the level of *Escherichia* were removed from the dataset. Due to the differences in diversity seen between adults and children, these datasets were analysed independently. Low abundance taxa (< 20 reads per sample) were removed and proportions of total bacterial reads calculated for remaining taxa. Taxa with proportions below 0.005 for all samples were dropped from the analysis. For comparative analysis, the log₂ median ratio of taxon abundances was calculated for each pairwise comparison. Statistical significance between groups was measured using the Wilcoxon rank sum test, with an adjustment for false discovery rate. For visual heat tree comparisons, log₂ median ratios were filtered to display only those that were significantly different between groups.

Results

Burden of *E. coli* in symptomatically and asymptotically ETEC-infected individuals

ETEC culture positive stool samples were collected from 32 symptomatic and asymptomatic adults and children at the icddr'b in Dhaka, Bangladesh. As controls, 16 ETEC culture negative stool samples were collected from healthy adults and children in the same city. DNA was extracted from stool samples and shotgun sequenced by Illumina HiSeq. To determine the approximate ETEC load in each sample, sequenced reads were mapped to the O78 ETEC reference genome H10407. The percent of reads that mapped to *E. coli* H10407 was then compared for each health status and age group.

Regardless of age, participants who were symptomatically infected with ETEC were more likely to have high percentages of reads mapping to *E. coli* H10407 (Figure 1A). The average percent of reads that mapped to *E. coli* H10407 was also significantly greater in symptomatic participants compared to asymptotically infected participants and ETEC-negative healthy controls. Asymptomatic participants had a higher average percent of *E. coli* H10407 reads when compared to controls ($p = 0.002$) but were significantly lower than the symptomatic group. While this suggests that symptomatically infected participants had a higher burden of ETEC on average, it is possible that they may also have had an abundance of other *E. coli* or related bacteria, or that they were pre-symptomatic and would go on to have diarrheal symptoms.

To specifically identify ETEC and other diarrhoeagenic *E. coli* in the stool samples, pathotype defining virulence gene sequences were detected in shotgun data using SRST2. ETEC positive samples were defined as those where we could detect DNA sequences for either the heat-labile toxin (LT) or heat-stable toxin (ST) genes. Of the 15 symptomatic participants in the study, we could detect LT or ST in 14 samples (Figure 1B). Importantly, the proportion of samples within which we could detect LT/ST and the reads per kilobase mapped (RPKM) for these genes was significantly greater in the symptomatically infected group when compared to asymptotically infected individuals (Wilcoxon rank-sum test, LT $p = 0.0285$; ST $p = 0.0413$). There was no significant difference in the detection of reads between adults and children.

In addition to ETEC-associated DNA, symptomatically infected participants had significantly higher levels of detectable DNA associated with enteroaggrative *E. coli* (EAEC) than asymptotically infected participants ($p = 0.003$). DNA sequences normally associated with enteropathogenic *E. coli* (EPEC) were also detected in some samples, but there was no statistically significant difference between groups (Figure 1B). When compared to control participants, symptomatically infected individuals had significantly higher detection of ETEC and EAEC associated DNA, while for asymptomatic participants, only ETEC-associated DNA signatures were significantly more prevalent ($p = 0.009$).

Interestingly, of the samples from which we could detect ETEC-associated DNA, the majority of symptomatic patients were infected with ETEC harbouring genes for both LT and ST. In contrast, the most common ETEC found in asymptomatic patients was ST-only ETEC. ETEC positive for the LT gene only were only found in one patient in each infected group. Of the ETEC colonization factor adhesins identified, CFA and CS13 were significantly more prevalent in asymptomatic and symptomatic participants than controls (Figure 1B). A number of additional adhesins were detected in symptomatic participants, including CS1, CS2, CS6, CS14, CS17 and CS22. Overall, the number of fimbriae detected per sample was higher in symptomatic participants compared to those who were asymptomatic (Mann Whitney u test, $p < 0.01$). This could be due to mixed ETEC infections, or a higher average number of fimbriae produced by ETEC strains colonising participants in the symptomatic group.

Comparison of microbiome diversity and composition between asymptomatic and symptomatic individuals

To compare the gut microbiome composition of stool samples from ETEC infected participants and controls, human sequence-depleted DNA reads were assigned informatically into taxonomic groups and the microbial composition was compared at the Phylum level (Figure 2). As expected, there were noticeable differences between the microbial composition of stool between adults and children. In general, healthy adults were heavily colonised by Firmicutes, while healthy children were predominantly colonised by Actinobacteria. In contrast, symptomatic participants showed high proportions of Proteobacteria, regardless of age. While the majority of asymptomatic participants were similar to controls at the phylum level, there were some notable exceptions within the dataset where asymptomatic participants looked more similar to symptomatic participants than to controls.

The differences in microbiome structure were further investigated with additional granulation by determining the richness (a-diversity) and composition (b-diversity) of the samples at the genus level. Alpha diversity was measured using Shannon's Diversity Index, a measure of population abundance and evenness. Higher Shannon indices are seen in samples which are more diverse in the number of species present, and lower when the sample is populated by fewer species or predominantly composed of few species. As expected, the pattern of a-diversity was very different between adults and children (Figure 3). The a-diversity of the microbiome in control and asymptomatic adults was significantly greater than that of symptomatic participants (Figure 3A). In children however, the microbiome a-diversity in controls was found to be much lower than adult controls (Figure 3A). There was a trend towards an increase in a-

diversity in asymptomatic children when compared to controls, however this difference was not statistically significant. Symptomatic children had diversity profiles somewhere between controls and asymptomatic children, but were not significantly different to either.

Beta diversity was compared by calculating the Bray Curtis Dissimilarity Index, which compares the compositional dissimilarity between samples. A Bray Curtis value of 0 reflects a situation whereby samples contained exactly the same microbes, whereas a value of 1 would suggest they did not share any microbial species. In general, symptomatically infected adults and children had similar microbial composition to each other, regardless of age (Figure 3B, dark green and dark brown ellipses). Control adults and children had distinct microbial composition, as shown by the separate b-diversity populations (Figure 3B, light green and light brown ellipses). Asymptomatically infected participants did not cluster together, but were spread between the distinct control and symptomatic populations.

Diversity of antimicrobial resistance genes between asymptotically and symptomatically infected individuals and controls

Antimicrobial resistance genes were identified from whole shotgun sequence data using SRST2 and the ARGannot database. To allow comparison between samples, the relative abundance of each gene in the shotgun data was determined by calculating the reads per kilobase mapped (RPKM) (Figure 4). There were significant differences in the antibiotic genes detected in the different health status groups. In general, genes of the beta lactamase, fluoroquinolone, sulfonamide and trimethoprim classes were more common in symptomatic and asymptomatic patients than controls. Only *aadA* and *tetB* were significantly more prevalent in symptomatic patients than in asymptomatic patients. Conversely, *ACI-1* and *ermQ* were both found to be more prevalent in asymptomatic patients than controls. There were few remarkable differences between the age groups, however in control participants a small number of tetracycline genes (*tet-40* and *tetB-P*) were more prevalent in adults than children.

Differentially abundant taxa between individuals with high and low *E. coli* loads

The above analysis of the microbiome samples showed significant heterogeneity between asymptotically infected participants. It is clear however that despite ETEC exposure, some individuals were able to protect themselves from an outgrowth of *E. coli*, while others became heavily colonised. To further explore the differences in the microbiomes of people who were highly colonised with ETEC compared to those with low or no ETEC colonisation, we reclassified samples into two categories - those with high *E. coli* burden (> 10% reads mapped to *E. coli* H10407) and those with low *E. coli* burdens (< 10% reads mapped to *E. coli* H10407). For adults, 9/25 (36%) of enrolled participants had high proportions of *E. coli* reads, while for children this was 10/23 (43%).

To allow us to identify underlying differences in the non-*E. coli* bacterial populations in participants infected with ETEC, the reads classified as Genus *Escherichia* were removed from the dataset before analysis. For comparison between groups, abundance data at the genus level was normalised, and pairwise comparisons between each group were made by calculating the log₂ ratio of median

proportions. The median ratios were then filtered to remove hits that were not significantly different between groups. The significant differences between the high and low *E. coli* burden groups were then visualised using heat tree maps (Figure 5). To provide additional granularity, this analysis was repeated at the species level (Supplementary Tables 1A and 1B).

There were 101 and 43 species significantly different in abundance in adults and children, respectively. Adults and children with high *E. coli* burdens had higher proportions of numerous other Enterobacteriaceae compared to those with low *E. coli* burdens. At the species level this included *Salmonella enterica*, *Citrobacter* spp and *Klebsiella* spp. Adults with high *E. coli* reads also harboured higher proportions of reads that mapped to the genomes of Campylobacteria, Burkholderiales, *Pasteurellaceae* and undefined *Veillonellaceae* F0422 MGYG-HGUT-03197 (Figure 5A). In contrast, adults with low *E. coli* burden had significantly higher proportions of numerous as yet undefined species, including members of the classes Bacilli and Clostridia.

For children, there were fewer statistically significant differences between the two groups, and fewer taxa that contributed greater than 0.005 of reads to the population, proportionally (Figure 5B). In addition to members of the *Enterobacteriaceae*, other species that were highly enriched in children with high *E. coli* reads were *Streptococcus* spp., *Comamonas* spp. and the newly defined species F0422 MGYG-HGUT-03197, which was also identified in adults. The only taxa that were significantly enriched in children with low *E. coli* burden were six *Bifidobacterium* species (*infantis*, *scardovii*, *pullorum*, *gallinarum*, *breve* and *angulatum*).

Discussion

In this study, we found significant differences in microbial diversity among symptomatically and asymptotically ETEC infected individuals compared to controls in both adults and children. The most striking difference between symptomatic and asymptomatic participants was the number of reads mapping to the widely used reference ETEC H10407 genome. Although our analysis may not provide an exact measure of ETEC burden, as it includes other *E. coli*, the *E. coli* read depths largely correlated with detected LT and ST toxin genes, which are key signatures of ETEC. This strongly suggests that asymptomatic people have a lower ETEC burden than symptomatic individuals. In this proof of concept study, asymptomatic participants were substantially more likely to harbour ST-only ETEC, whereas more symptomatic participants had LT/ST-ETEC. However, why ETEC are able to persist and replicate in some individuals without causing clinical disease remains an open question. While the presence of LT could increase the likelihood of a patient becoming symptomatic, data from large scale studies suggests that ST is the more important toxin for mediating moderate-to-severe diarrhoea (2).

One limitation of this small, proof-of-concept study was that samples were only collected from participants at one timepoint, and participants with detectable ETEC in their stool were not followed up to establish whether or not a subset would eventually go on to become symptomatic. While progression to symptomatic infection can not be discounted within this cohort, we do know from case-control studies

such as GEMS1 and GEMS1A that ETEC (LT or ST positive) is commonly identified in asymptomatic individuals (2,24). From the GEMS1 study in Bangladesh, only ST-positive ETEC were significantly associated with disease, although this was only in the younger age groups. In children between 24 – 59 months, all cause ETEC (ST or LT positive) were detected in 20 cases and 26 controls (<https://clinepidb.org/ce/app/>). Thus, there is a clear case for asymptomatic carriage of ETEC in children and adults. Building on this work, a larger study collecting longitudinal samples from participants will be needed to investigate how the microbiome changes in response to ETEC colonisation over time and what factors precipitate progression to symptomatic disease.

There were no significant differences between adults and children in the *E. coli* read depth, ETEC toxin gene carriage, ARG carriage or the presence of other *E. coli* pathotypes. However, the microbiome diversity and response to infection was considerably different between these two groups. Microbial diversity in control children was relatively low, and their microbiomes were dominated by Actinobacteria. The diversity in control children was distinct from asymptotically infected children, who had much higher α -diversity. This likely represents the proliferation of species normally present at low abundances in these children due to disruption of normal gut homeostasis. Whether ETEC infection precipitates this change or is a result of it is yet to be determined. In general, when looking at microbial diversity the microbiomes of symptomatic and asymptomatic children were more similar to each other than to controls. This is an important factor to consider when deciding on appropriate controls for future microbiome research or potential microbial therapeutic development.

On the species level we also saw increases in the proportions of several *Streptococcal* species in children with high reads. This is consistent with results of a similar study on microbiome changes during intestinal infection in Bangladesh, which found proportional increases in *E. coli* and several *Streptococcal* species in children with diarrhea, regardless of the culpable intestinal pathogen (6). Conversely, in children with low *E. coli* burden we found higher proportions of several species of *Bifidobacterium*. These bacteria are common in breastfed infants (25), and were also increased in abundance in healthy children compared to diarrhea cases in the above study (7).

For adults, the overall microbiome profile was very different to that in children. Asymptomatic adults and controls had very similar gut microbiomes, which suggests that the introduction of ETEC does not singularly cause any considerable microbiome perturbation, and that there may be other factors at play. The healthy adult microbiome was also much more diverse, indicating decreased diversity upon ETEC infection as *E. coli* became the most abundant bacterial member of the microbiome. The microbiome in healthy adults was rich in Firmicutes, similar to other published studies from Bangladesh (14), although we also observed an abundance of Actinobacteria. This finding was similar to a study on travelers returning to the US from a LMIC, which found an abundance of Firmicutes, Actinobacteria and Proteobacteria in healthy and ETEC infected travelers (19). Although a study on ETEC infections in healthy volunteers in the US found a number of *Bacteroides* species to be protective against infection (14), in our study the *Bacteroides* genus was on average greater than 2-fold more abundant in adults with

high *E. coli* sequence reads, although no species were significantly more abundant. This may be related to baseline differences in gut microbiome composition between the US and Bangladesh.

It is clear from this data that progression to symptomatic ETEC infection is associated with an outgrowth of *Enterobacteriaceae*. Although interestingly, many uncharacterised species were also significantly associated with either protection from or progression towards symptomatic disease. Most of these bacteria have only recently been discovered, and many have never been cultured as they were identified through metagenome assembly. This highlights the importance of efforts like the Unified Human Gastrointestinal Genome to shine a light on previously hidden microbial populations (19).

Finally, ARG carriage was significantly different between the groups, with asymptomatic and symptomatic patients being broadly more similar to each other than controls. In general, patients colonised with ETEC had higher prevalence of ARGs of the trimethoprim, sulfonamide, fluoroquinolone and β -lactamase classes. These antibiotics are commonly used in South Asia, both clinically and in farming, and high levels of resistance to these classes has been seen in other studies (26–29). Although there was increased prevalence of ARGs in people colonised with ETEC, it is unlikely that all the ARGs would be ETEC-associated, as most *E. coli* isolates in South Asia are not so highly multidrug resistant (26,27). It is therefore probable that the increase in ARG prevalence is associated with changes in the microbiome and increases in the relative proportions of ARG carrying *Enterobacteriaceae*. Understanding the mechanics of this interaction will require larger studies encompassing longitudinal sampling and comprehensive lifestyle and antibiotic usage surveys.

Conclusions

Our study highlights the gastrointestinal microbiome differences between asymptomatic colonisation with ETEC, compared to symptomatic infection. Not only were there clear alterations in microbial composition in these patients, but the prevalence of several clinically relevant ARG classes was also impacted. This work provides important foundational information for future studies investigating the impact of gut microbiome composition on gastrointestinal disease progression, and how the presence of supportive organisms may endow protection against symptomatic disease.

Declarations

Ethics approval and consent to participate

The samples and data collected for this study were collected as part of the ETEC ETVAX Vaccine Trial (clinical trial NCT02531802).

Consent for publication

Not applicable

Availability of data and material

The sequence data is available on the European Nucleotide Archive under Study Accession PRJEB21793.

Competing interests

The authors declare they have no competing interest.

Funding

At the University of Cambridge, this work was funded by the NIHR AMR Research Capital Funding Scheme [NIHR 200640]. This work was also supported by the icddr,b, which is grateful to the governments of Bangladesh, Canada, Sweden and the UK for providing core/unrestricted support. GD and AM were supported by NIHR-BRC. AM was also funded by the DBT-Cambridge partnership. Additional support was provided by Wellcome Trust Strategic Award (no. 106158/Z/14/Z) and the Bill and Melinda Gates Foundation (no. 617 OPP1141321).

Authors contributions

EH – investigation, data analysis, writing-original draft and editing.

MAS – investigation, writing-review and editing.

JPD – data analysis, writing-review and editing.

VS – writing-review and editing.

MB – writing-review and editing.

SKS – writing-review and editing.

FQ – conceptualisation, funding, writing-review and editing.

GD – conceptualisation, supervision, funding, writing-review and editing.

AM – conceptualisation, supervision, funding, writing-review and editing

Acknowledgements

The authors would like to thank the clinical and laboratory teams at the International Centre for Diarrheal Disease Research, Bangladesh, and the sequencing team at the Wellcome Sanger Institute for their assistance.

References

1. Anderson IV JD, Bagamian KH, Muhib F, Amaya MP, Laytner LA, Wierzba T, et al. Articles Burden of enterotoxigenic *Escherichia coli* and shigella non-fatal diarrhoeal infections in 79 low-income and lower middle-income countries: a modelling analysis [Internet]. 2019 [cited 2021 Jan 28]. Available from: www.thelancet.com/lancetgh
2. Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, Panchalingam S, et al. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): A prospective, case-control study. *The Lancet*. 2013;382(9888):209–22.
3. de Filippo C, Cavalieri D, di Paola M, Ramazzotti M, Baptiste Poullet J, Massart S, et al. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. [cited 2021 Jan 28]; Available from: www.pnas.org/cgi/doi/10.1073/pnas.1005963107
4. Yatsunencko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. Vol. 486, *Nature*. 2012. p. 222–7.
5. Bhute S, Pande P, Shetty SA, Shelar R, Mane S, Kumbhare S v., et al. Molecular Characterization and Meta-Analysis of Gut Microbial Communities Illustrate Enrichment of *Prevotella* and *Megasphaera* in Indian Subjects. *Frontiers in Microbiology* [Internet]. 2016 May 9 [cited 2021 Jan 28];7(MAY):660. Available from: <http://journal.frontiersin.org/Article/10.3389/fmicb.2016.00660/abstract>
6. Lin A, Bik EM, Costello EK, Dethlefsen L, Haque R, Relman DA, et al. Distinct Distal Gut Microbiome Diversity and Composition in Healthy Children from Bangladesh and the United States. Aziz RK, editor. *PLoS ONE* [Internet]. 2013 Jan 22 [cited 2019 Dec 9];8(1):e53838. Available from: <http://dx.plos.org/10.1371/journal.pone.0053838>
7. Youmans BP, Ajami NJ, Jiang Z-D, Campbell F, Wadsworth WD, Petrosino JF, et al. Characterization of the human gut microbiome during travelers' diarrhea. *Gut Microbes* [Internet]. 2015 Mar 4 [cited 2019 Dec 9];6(2):110–9. Available from: <https://www.tandfonline.com/doi/full/10.1080/19490976.2015.1019693>
8. Midani FS, Weil AA, Chowdhury F, Begum YA, Khan AI, Debela MD, et al. Human Gut Microbiota Predicts Susceptibility to *Vibrio cholerae* Infection. *The Journal of Infectious Diseases* [Internet]. 2018 Jul 13 [cited 2019 Dec 19];218(4):645–53. Available from: <https://academic.oup.com/jid/article/218/4/645/4969495>
9. Lindsay B, Oundo J, Hossain MA, Antonio M, Tamboura B, Walker AW, et al. Microbiota that affect risk for shigellosis in children in low-income countries. *Emerging Infectious Diseases*. 2015;21(2):242–50.
10. Weil A, Midani F, Chowdhury F, Khan A, Begum Y, Charles R, et al. The Gut Microbiome and Susceptibility to *Vibrio cholerae* Infection. *Open Forum Infectious Diseases* [Internet]. 2015 Dec 9 [cited 2019 Dec 19];2(suppl_1). Available from: <https://academic.oup.com/ofid/article/doi/10.1093/ofid/ofv131.72/2633983>

11. Kieser S, Sarker SA, Sakwinska O, Foata F, Sultana S, Khan Z, et al. Bangladeshi children with acute diarrhoea show faecal microbiomes with increased *Streptococcus* abundance, irrespective of diarrhoea aetiology. *Environmental Microbiology* [Internet]. 2018 Jun 7 [cited 2019 Dec 16];20(6):2256–69. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/1462-2920.14274>
12. Monira S, Nakamura S, Gotoh K, Izutsu K, Watanabe H, Alam NH, et al. Metagenomic profile of gut microbiota in children during cholera and recovery. *Gut Pathogens* [Internet]. 2013 [cited 2019 Dec 9];5(1):1. Available from: <http://gutpathogens.biomedcentral.com/articles/10.1186/1757-4749-5-1>
13. Singh P, Teal TK, Marsh TL, Tiedje JM, Mosci R, Jernigan K, et al. Intestinal microbial communities associated with acute enteric infections and disease recovery. *Microbiome* [Internet]. 2015 Dec 22 [cited 2019 Nov 20];3(1):45. Available from: <http://www.microbiomejournal.com/content/3/1/45>
14. Pop M, Paulson JN, Chakraborty S, Astrovskaya I, Lindsay BR, Li S, et al. Individual-specific changes in the human gut microbiota after challenge with enterotoxigenic *Escherichia coli* and subsequent ciprofloxacin treatment. *BMC Genomics* [Internet]. 2016 Dec 8 [cited 2019 Dec 9];17(1):440. Available from: <http://bmcbgenomics.biomedcentral.com/articles/10.1186/s12864-016-2777-0>
15. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature Methods*. 2012 Apr;9(4):357–9.
16. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, et al. VFDB: A reference database for bacterial virulence factors. *Nucleic Acids Research*. 2005 Jan 1;33(DATABASE ISS.).
17. Inouye M, Dashnow H, Raven L-A, Schultz MB, Pope BJ, Tomita T, et al. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Medicine* [Internet]. 2014 Dec 20 [cited 2019 Nov 7];6(11):90. Available from: <http://genomemedicine.biomedcentral.com/articles/10.1186/s13073-014-0090-6>
18. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *bioRxiv* [Internet]. 2019 [cited 2019 Nov 7];762302. Available from: <https://www.biorxiv.org/content/10.1101/762302v1>
19. Almeida A, Nayfach S, Boland M, Strozzi F, Beracochea M, Shi ZJ, et al. A unified sequence catalogue of over 280,000 genomes obtained from the human gut microbiome. *bioRxiv*. 2019 Sep 19;762682.
20. Lu J, Breitwieser F, Thielen P, Science SS-PC, 2017 undefined. Bracken: estimating species abundance in metagenomics data. *peerj.com* [Internet]. [cited 2019 Nov 7]; Available from: <https://peerj.com/articles/cs-104/>
21. Oksanen J. *Vegan: ecological diversity*.
22. de Mendiburu F. Package “agricolae.” 2019.

23. Foster ZSL, Sharpton TJ, Grünwald NJ. Metacoder: An R package for visualization and manipulation of community taxonomic diversity data. Poisot T, editor. *PLOS Computational Biology* [Internet]. 2017 Feb 21 [cited 2019 Nov 7];13(2):e1005404. Available from: <https://dx.plos.org/10.1371/journal.pcbi.1005404>
24. Kotloff KL, Nasrin D, Blackwelder WC, Wu Y, Farag T, Panchalingham S, et al. The incidence, aetiology, and adverse clinical consequences of less severe diarrhoeal episodes among infants and children residing in low-income and middle-income countries: a 12-month case-control study as a follow-on to the Global Enteric Multicenter Study (GEMS). *The Lancet Global Health* [Internet]. 2019 May 1 [cited 2021 Jan 29];7(5):e568–84. Available from: www.thelancet.com/lancetgh
25. Bezirtzoglou E, Tsiotsias A, Welling GW. Microbiota profile in feces of breast- and formula-fed newborns by using fluorescence in situ hybridization (FISH). *Anaerobe*. 2011 Dec;17(6):478–82.
26. Talukdar PK, Rahman M, Rahman M, Nabi A, Islam Z, Hoque MM, et al. Antimicrobial Resistance, Virulence Factors and Genetic Diversity of *Escherichia coli* Isolates from Household Water Supply in Dhaka, Bangladesh. *PLoS ONE* [Internet]. 2013 Apr 3 [cited 2021 May 26];8(4):e61090. Available from: www.plosone.org
27. Natarajan M, Kumar D, Mandal J, Biswal N, Stephen S. A study of virulence and antimicrobial resistance pattern in diarrhoeagenic *Escherichia coli* isolated from diarrhoeal stool specimens from children and adults in a tertiary hospital, Puducherry, India. *Journal of Health, Population and Nutrition* [Internet]. 2018 Jul 13 [cited 2021 May 25];37(1):1–11. Available from: <https://doi.org/10.1186/s41043-018-0147-z>
28. Swarouth JM, Fuhrmeister ER, Hamzah L, Harris AR, Ahmed MA, Gurley ES, et al. Differential overlap in human and animal fecal microbiomes and resistomes in rural versus urban Bangladesh. [cited 2021 May 25]; Available from: <https://doi.org/10.1101/2021.05.13.21257188>
29. Ahmed I, Rabbi MB, Sultana S. Antibiotic resistance in Bangladesh: A systematic review. Vol. 80, *International Journal of Infectious Diseases*. Elsevier B.V.; 2019. p. 54–61.

Tables

Table 1. Study participants by age and health status

	Control	Asymptomatic	Symptomatic	Total
Children	8	7	8	23
Adults	8	10	7	25
Total	16	17	15	48

Figures

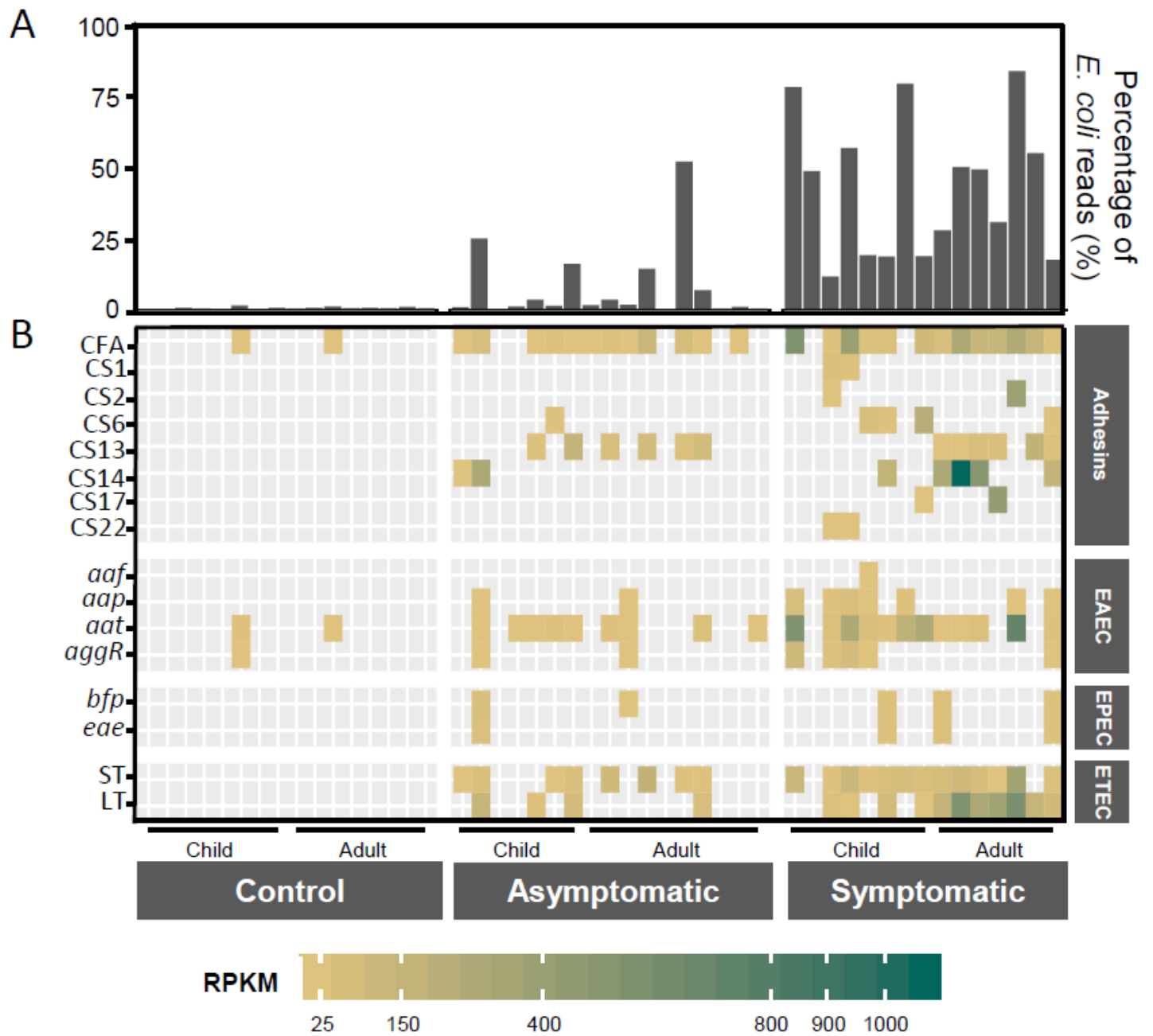


Figure 1

Detection of *E. coli* and associated virulence factors from sequence data. For each sample the sequence data was mined to determine the percent of reads mapping to *E. coli* H10407 (A) and the presence of adhesins and pathogenic *E. coli* associated virulence factors (B). The intensity of the heatmap is shown relative to the reads per kilobase mapped (RPKM) for the gene of interest.

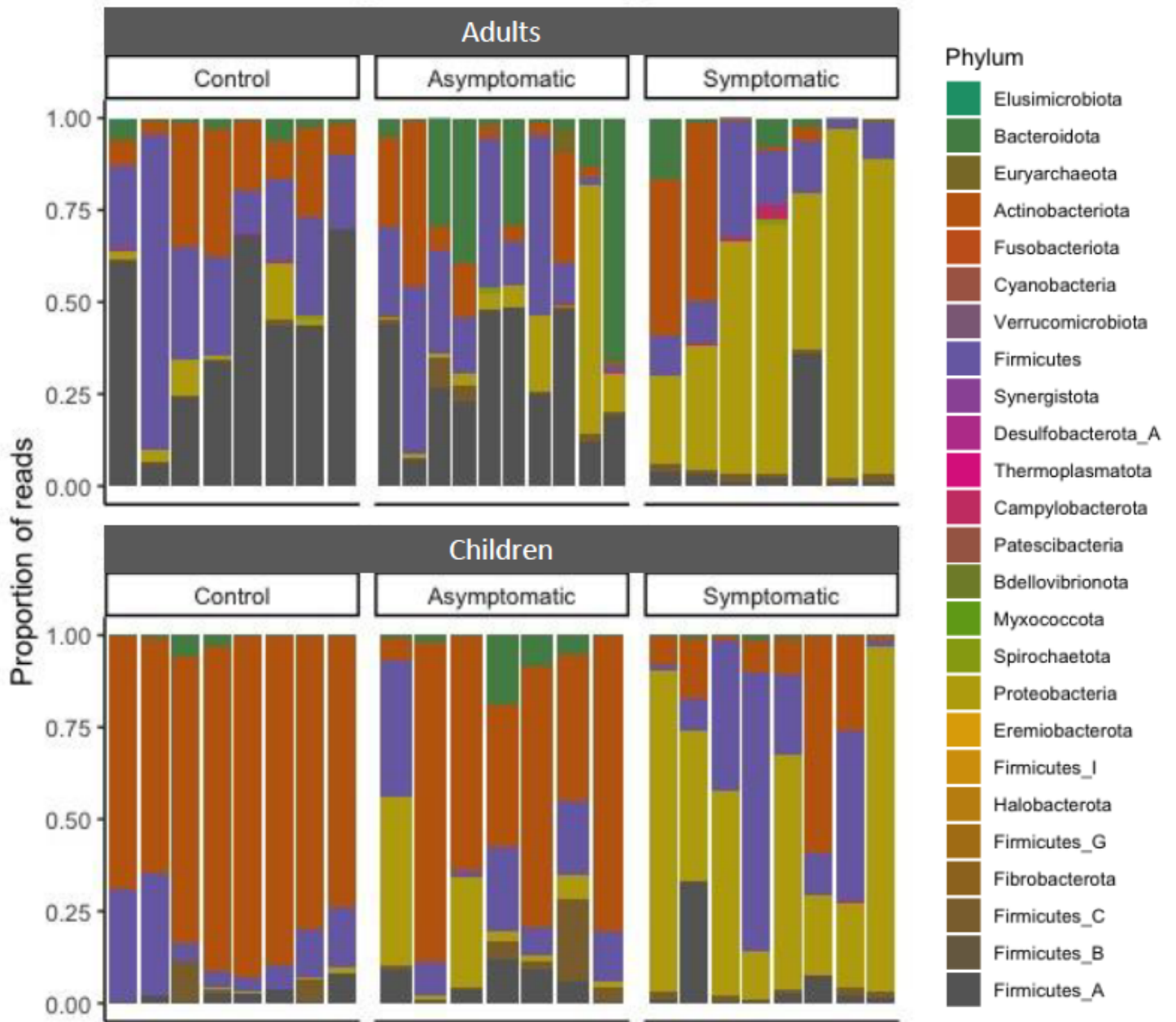


Figure 2

Microbial composition of gut microbiome samples by age group and health status. Sequenced reads were classified to the phylum level, and the relative abundance of each phylum is shown as the proportion of total bacterial reads.

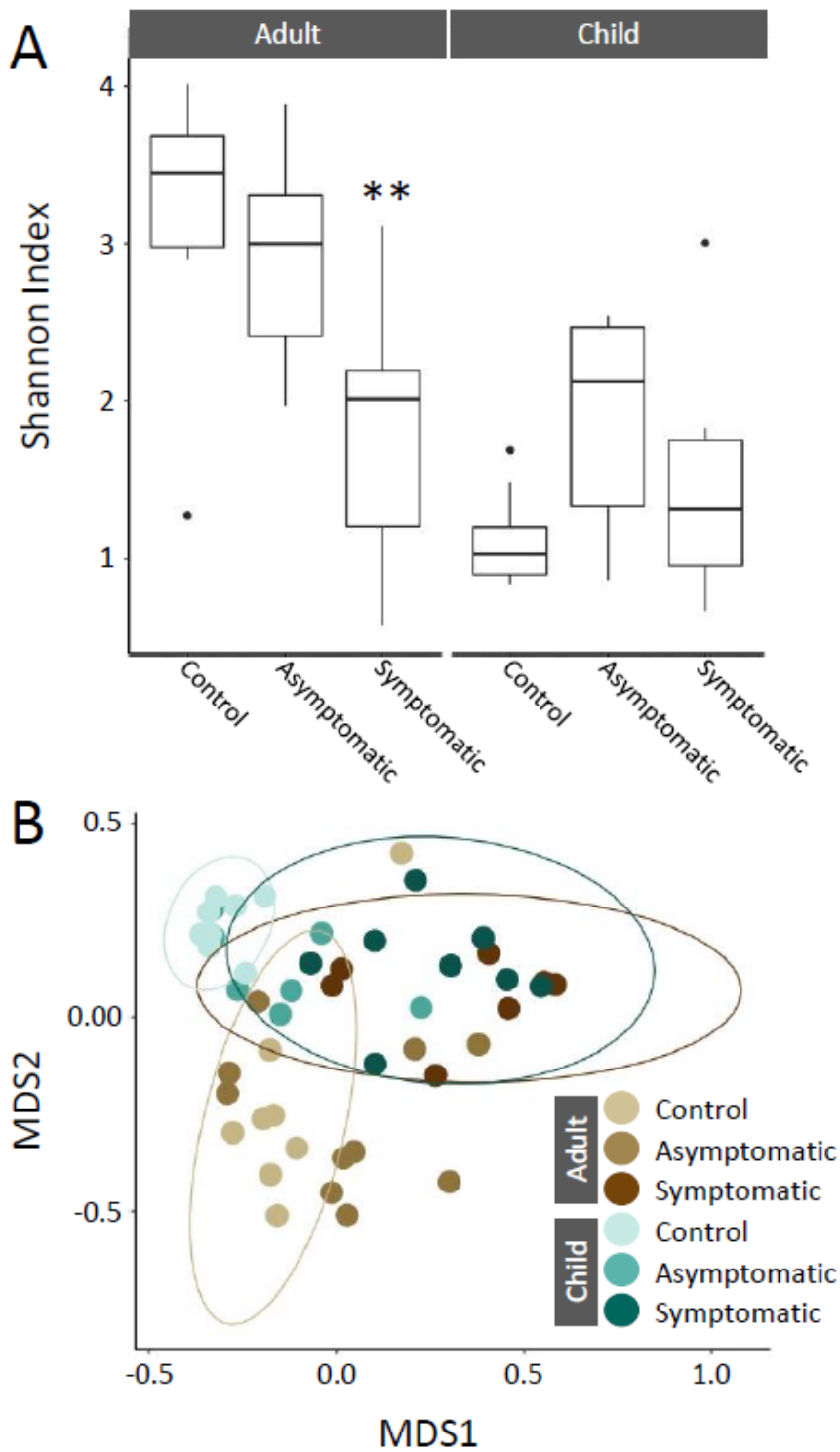


Figure 3

Diversity of stool microbiome samples by health status group and age. The diversity of microbiome samples was compared at the genus level and stratified by age group and health status. (A) Shannon (H') diversity (B) Bray Curtis Dissimilarity (β diversity). ** $p < 0.01$.

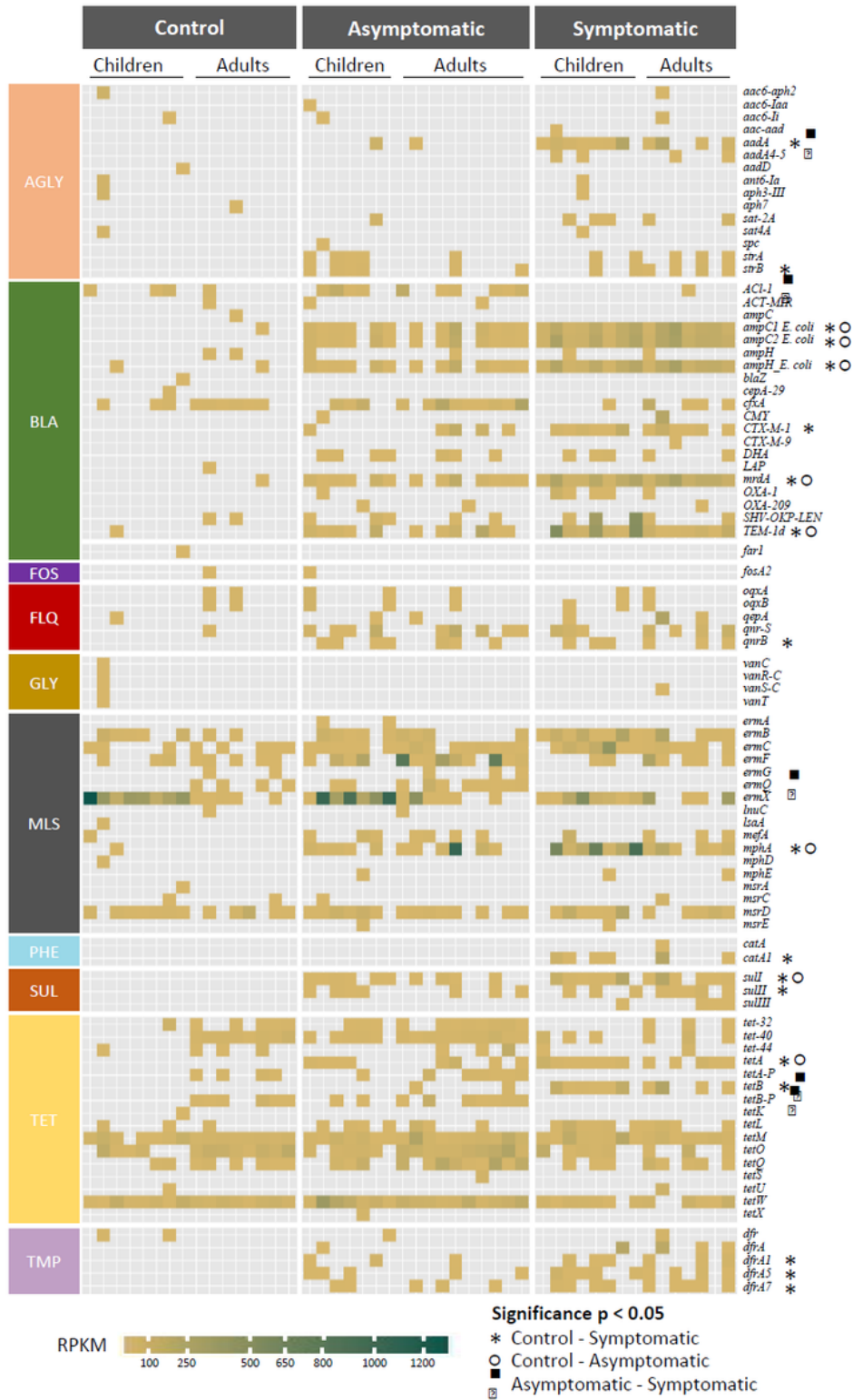


Figure 4

Antimicrobial resistance gene carriage in participants. Reads per kilobase mapped (RPKM) for ARGs detected in participant stool samples are shown as a heatmap, clustered by health status, age and antimicrobial class. Significant differences (Fishers exact test) in the prevalence of specific ARGs between groups is shown next to the gene name. AGLY – aminoglycosides, BLA – beta lactamases, FOS – fosfomycin, FLQ – fluoroquinolones, GLY – glycosides, MLS –

macrolides/lincosamides/streptogramins, PHE – phenicols, SUL – sulfonamides, TET – tetracyclines, TMP – trimethoprim.

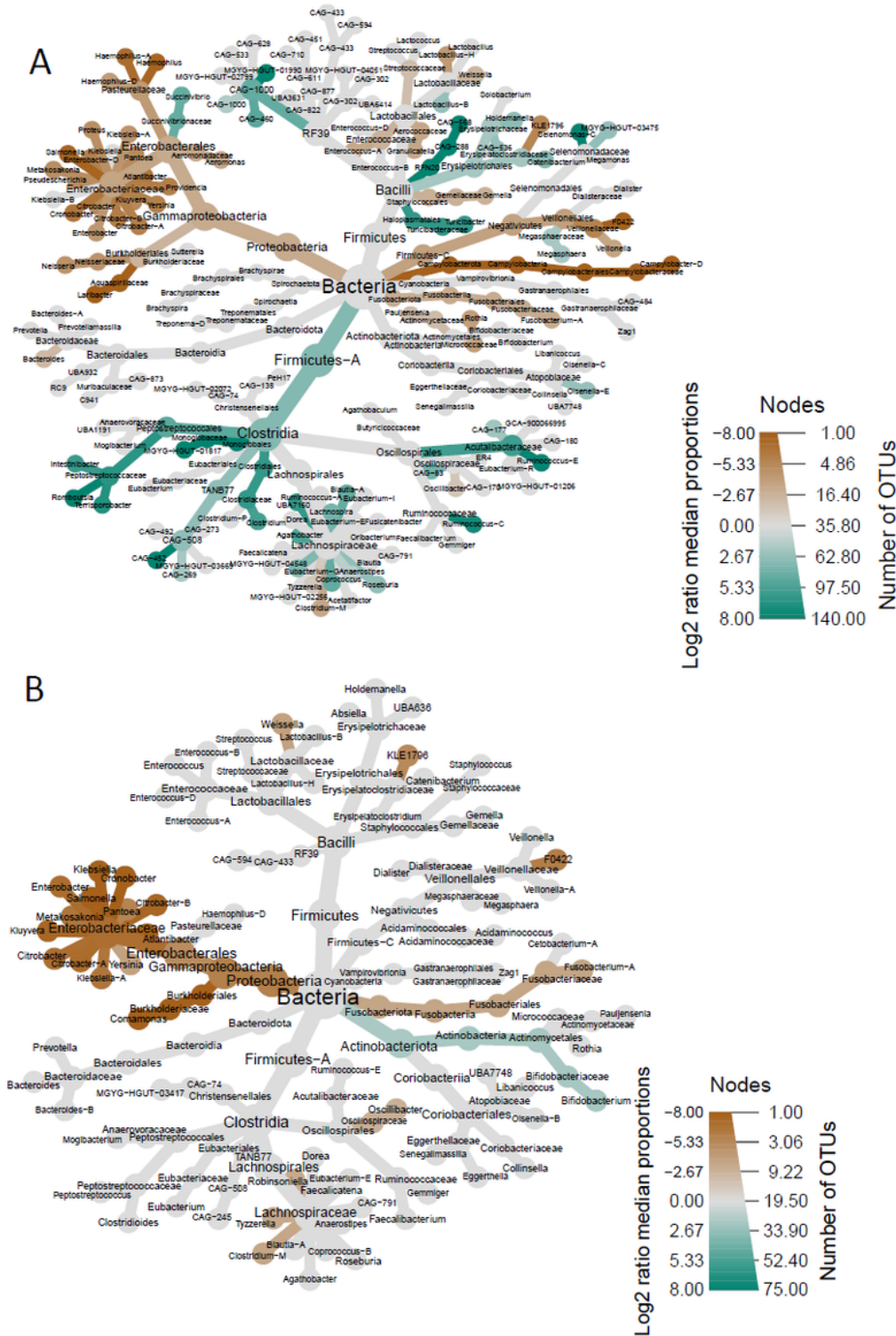


Figure 5

Heat trees showing statistically significant differences in abundance of bacterial taxa in adults (A) and children (B) with high (brown) or low (green) *E. coli* burden. The size of the node is proportional to the number of OTUs detected, while the depth of colour is proportional to the size of the log₂ median

difference between groups. Grey branches show taxa that were present, but not significantly different between groups.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.xlsx](#)