

RESEARCH ARTICLE

Meta-taxonomic analysis of prokaryotic and eukaryotic gut flora in stool samples from visceral leishmaniasis cases and endemic controls in Bihar State India

Rachael Lappan¹, Cajsa Classon², Shashi Kumar³, Om Prakash Singh³, Ricardo V. de Almeida⁴, Jaya Chakravarty³, Poonam Kumari³, Sangeeta Kansal³, Shyam Sundar^{3‡}, Jenefer M. Blackwell^{1,5‡*}

1 Telethon Kids Institute, The University of Western Australia, Nedlands, Western Australia, Australia, **2** Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Solna, Sweden, **3** Department of Medicine, Institute of Medical Sciences, Banaras Hindu University, Varanasi, India, **4** Departamento de Bioquímica, Centro de Biociências, Universidade Federal do Rio Grande do Norte, Natal, Brazil, **5** Department of Pathology, University of Cambridge, Cambridge, United Kingdom

‡ These authors are joint senior authors on this work.

* jenefer.blackwell@telethonkids.org.au



OPEN ACCESS

Citation: Lappan R, Classon C, Kumar S, Singh OP, de Almeida RV, Chakravarty J, et al. (2019) Meta-taxonomic analysis of prokaryotic and eukaryotic gut flora in stool samples from visceral leishmaniasis cases and endemic controls in Bihar State India. *PLoS Negl Trop Dis* 13(9): e0007444. <https://doi.org/10.1371/journal.pntd.0007444>

Editor: Guy Caljon, Universiteit Antwerpen, BELGIUM

Received: May 1, 2019

Accepted: August 12, 2019

Published: September 6, 2019

Copyright: © 2019 Lappan et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The dataset generated and analysed during the current study is available in the NCBI Sequence Read Archive (PRJNA525566; available at <https://www.ncbi.nlm.nih.gov/sra/PRJNA525566>). Complete documentation of all 16S rRNA and 18S rRNA sequence analysis code can be found at <https://rachaellappan.github.io/VL-QIIME2-analysis/>.

Funding: This work was supported by the NIH as part of Tropical Medicine Research Centre award

Abstract

Visceral leishmaniasis (VL) caused by *Leishmania donovani* remains of public health concern in rural India. Those at risk of VL are also at risk of other neglected tropical diseases (NTDs) including soil transmitted helminths. Intestinal helminths are potent regulators of host immune responses sometimes mediated through cross-talk with gut microbiota. We evaluate a meta-taxonomic approach to determine the composition of prokaryotic and eukaryotic gut microflora using amplicon-based sequencing of 16S ribosomal RNA (16S rRNA) and 18S rRNA gene regions. The most abundant bacterial taxa identified in faecal samples from Bihar State India were *Prevotella* (37.1%), *Faecalibacterium* (11.3%), *Escherichia-Shigella* (9.1%), *Alloprevotella* (4.5%), *Bacteroides* (4.1%), *Ruminococcaceae* UCG-002 (1.6%), and *Bifidobacterium* (1.5%). Eukaryotic taxa identified (excluding plant genera) included *Blastocystis* (57.9%; Order: Stramenopiles), *Dientamoeba* (12.1%; Family: Tritrichomonadea), *Pentatrichomonas* (10.1%; Family: Trichomonodea), *Entamoeba* (3.5%; Family: Entamoebida), Ascaridida (0.8%; Family: Chromodorea; concordant with *Ascaris* by microscopy), Rhabditida (0.8%; Family: Chromodorea; concordant with *Strongyloides*), and Cyclophyllidea (0.2%; Order: Eucestoda; concordant with *Hymenolepis*). Overall alpha (Shannon's, Faith's and Pielou's indices) and beta (Bray-Curtis dissimilarity statistic; weighted UniFrac distances) diversity of taxa did not differ significantly by age, sex, geographic subdistrict, or VL case (N = 23) versus endemic control (EC; N = 23) status. However, taxon-specific associations occurred: (i) *Ruminococcaceae* UCG-014 and *Gastranaerophilales_ uncultured bacterium* were enriched in EC compared to VL cases; (ii) *Pentatrichomonas* was more abundant in VL cases than in EC, whereas the reverse occurred for *Entamoeba*. Across the cohort, high *Escherichia-Shigella* was associated with reduced bacterial diversity, while high *Blastocystis* was associated with high bacterial

U19 AI074321 (SS and JB), The Swedish Research Council Vetenskapsrådet (VR) grant no 348-2014-2819 CC), and by a travel grant to CC from The World Infection Fund (Världsinfectionsfonden (Vif). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

diversity and low *Escherichia-Shigella*. Individuals with high *Blastocystis* had low Bacteroidaceae and high *Clostridiales* vadin BB60 whereas the reverse held true for low *Blastocystis*. This scoping study provides useful baseline data upon which to develop a broader analysis of pathogenic enteric microflora and their influence on gut microbial health and NTDs generally.

Author summary

Visceral leishmaniasis (VL), also known as kala-azar, is a potentially fatal disease caused by intracellular parasites of the *Leishmania donovani* complex. VL is a serious public health problem in rural India, causing high morbidity and mortality, as well as major costs to local and national health budgets. People at risk of VL are also at risk of other neglected tropical diseases (NTDs) including soil transmitted helminths (worms). Intestinal worms are potent regulators of host immune responses often mediated through cross-talk with gut bacteria. Here we have used a modern DNA sequencing approach to determine the composition of microbiota in stool samples from VL cases and endemic controls. This allows us to determine all bacteria, as well as all single-celled and multicellular organisms, that comprise the microorganisms in the gut in a single sequencing experiment from a single stool sample. In addition to providing valuable information concerning commensal and pathogenic gut micro-organisms prevalent in this region of India, we find some specific associations between single-celled gut pathogens and VL case status.

Introduction

Visceral leishmaniasis (VL), also known as kala-azar, is a potentially fatal disease caused by obligate intracellular parasites of the *Leishmania donovani* species complex. VL is a serious public health problem in indigenous and rural populations in India, causing high morbidity and mortality, as well as major costs to both local and national health budgets. The estimated annual global incidence of VL is 200,000 to 400,000, with up to 50,000 deaths annually occurring principally in India, Bangladesh, Sudan, South Sudan, Ethiopia and Brazil [1].

The outcome of infection with *L. donovani* depends on the host immune response [2], that in turn could be influenced by other factors, one of which could be co-infection with other pathogens. Recent interest has focused in particular on the influence that gut microbial composition has on health and disease, in particular bacteria such as Enterobacteriaceae which have been associated with systemic inflammation (reviewed [3]). In the case of VL, in particular, it has been proposed that systemic inflammation due to microbial translocation from the gut is part of disease pathogenesis [4, 5]. Often those at risk of VL are also at risk for other neglected tropical diseases (NTDs) such as infestation with soil transmitted helminths or infection with lymphatic filariasis or *Schistosoma mansoni*. Intestinal helminths are particularly potent regulators of their host's immune response and can ameliorate inflammatory diseases [6]. Few published studies are available on interaction between *L. donovani* infection and other NTDs. Hassan and coworkers [7] showed that, despite the development of a functional anti-*L. donovani* T helper 1 response, mice with established *S. mansoni* infections fail to

control *L. donovani* growth in the liver and spleen. Similarly, we recently demonstrated [8] that chronic infection with intestinal nematode *Heligmosomoides polygyrus* exacerbated secondary *L. donovani* infection in mice, resulting in higher parasite burdens in liver and spleen compared to worm free mice. This increased parasite load was accompanied by increased interleukin-4 and interleukin-10 transcription in spleens. In contrast, the presence of the filarial parasite *Brugia malayi* L3/adult worms inhibited progression of *L. donovani* infection associated with a T helper 1 response in a hamster model [9]. However, much may depend on the timing of infection with the other NTD and with *L. donovani*. A chronic helminth infection such as may be expected in areas endemic for VL may have a different effect compared to a newly established primary infection. In humans, the only co-infection studies reported to date are in relation to American tegumentary leishmaniasis [10–12]. In an initial study of 120 patients with cutaneous leishmaniasis caused by *L. braziliensis* [11] patients co-infected with helminths took longer to heal than helminth-free patients. However, in a follow up study of 90 patients there was no statistically significant benefit to treating patients with anti-helminthics [12]. In a more recent retrospective cohort of 109 *L. braziliensis* patients from Brazil [10] it was shown that patients with positive parasitological stool examinations had more mucosal lesions and more often had a poor response to therapy. Interestingly, these effects were observed for total intestinal helminths (based cumulatively on Ancylostomidae, *Ascaris lumbricoides*, *Strongyloides stercoralis*, *Trichuris trichuria*, *Enterobius vermicularis*, *S. mansoni*), and for *A. lumbricoides* alone or nematodes cumulatively (*A. lumbricoides*, *Str. stercoralis*, *T. trichuria*, *E. vermicularis*), but not for protozoan (based cumulatively on *Blastocystis hominis*, *Entamoeba coli*, *E. histolytica*, *Giardia lamblia*, *Endolimas nana*) infections.

One limiting factor for wide-ranging analysis of gut pathogen load has been the specialized and time-consuming processes involved in microscopic identification of parasites. While multiplex PCR for fecal DNA samples is facilitating larger scale studies for specific helminth and protozoan gut pathogens [13] this does not provide data on the potential influence of pathogenic species on gut microbes more broadly. For example, it was recently reported that the gut microbiome contributes to impaired immunity in pulmonary tuberculosis patients in India [14]. The presence of gut helminths is known to influence the gut microbiome [15, 16], and there is evidence [6] of important cross-talk between the gut microbiome and helminths in mediating host immunological effects. A meta-taxonomic approach to look at both commensal and pathogenic species in the gut would therefore be of great value.

Numerous studies have used high-throughput, massively parallel amplicon-based sequencing of 16S ribosomal RNA (16S rRNA) gene regions to study the prokaryotic composition of the gut microflora [17–19], including studies in India [14, 20–24]. Recent studies in mammalian hosts [25] have also assessed parasite diversity using amplicon-based high-throughput sequencing of 18S ribosomal RNA (18S rRNA) gene regions and classified sequence reads into multiple parasite groups. Comparison of results with standard methods including microscopic observation of helminths in the intestines, and PCR amplification/ sequencing of 18S rDNA from isolated single worms, suggests that this new technique is reliable and useful to investigate eukaryotic parasite diversity in fecal samples. However, this approach has not to our knowledge been used in humans. Here we present a scoping study that uses 16S rRNA and 18S rRNA meta-taxonomic analysis of fecal samples with the primary aims of determining the composition of gut prokaryotic and eukaryotic microflora in an area of India endemic for VL and whether this differs between VL cases and non-VL endemic controls. We compare microscopic determination of helminth infections with 18S rRNA data, and report on secondary aims of determining the potential influence of prokaryotic or eukaryotic taxa known to contain pathogenic species on gut microbial diversity *per se*.

Methods

Ethics statement

The enrolment of human subjects complies with the principles laid down in the Helsinki declaration. Institutional ethical approval (reference numbers: Dean/2017/EC.148) was obtained from the ethical review board of Banaras Hindu University (BHU), Varanasi, India. Informed written consent was obtained from each participant at the time of enrolment, or from their legal guardian if they were under 18 years old. Only patients who had not previously received treatment and who agreed to participate in the study were enrolled. All clinical treatment and follow-up records were maintained using standardised case report forms on an electronic server.

Study subjects

Samples were collected between May and June 2017 at the Kala-azar Medical Research Center (KAMRC), Muzaffarpur, Bihar, India. Active VL cases (N = 23) were diagnosed by experienced clinicians based on clinical signs, including fever (>2 weeks), splenomegaly, positive serology for recombinant antigen (r)-K39 and/or by microscopic demonstration of *Leishmania* amastigotes in splenic aspirate smears. Fecal samples (cf. below) were collected prior to treatment. Endemic controls (EC; N = 23) that did not have VL were attendants accompanying VL patients at KAMRC (N = 16) or from field locations nearby (N = 7). Basic demographic details (age and sex) on participants are provided in [Table 1](#). Metadata by individual are provided in [S1 Table](#).

Stool collection

Stool collection containers and toilet accessories (to avoid urine contamination; OM-AC1; DNA Genotek) were distributed to the study participants and the containers collected immediately after defecation, either that evening or when participants defecated the following morning and stored at 4°C until processing. Most subjects defecated in the morning and the containers were then returned and processed within 3–4 hours of storage at 4°C. In rare instances defecation may have occurred at time when the nursing staff was not immediately available. Importantly, there was no difference in the collection and processing of samples between VL subjects and controls. The sample was divided into two parts: 200 µg to OMNI-gene GUT tubes (OM-200; DNA Genotek), and approximately 1 gram was concentrated by the formalin-ethyl acetate sedimentation method and used for diagnostic microscopy.

Parasite microscopy

For microscopy, formalin-ethyl-acetate sedimentation was used to concentrate parasites following standard methodology [26]. Briefly, faeces were homogenized in 10 ml of 10% formalin

Table 1. Basic demographic details of study participants. Full metadata are provided in [S1 Table](#).

Group	N	Age Mean±SD	Age Range
VL cases	23	21.5±15.3	5–55
Males	7	16.7±15.5	5–50
Females	16	23.6±15.2	5–55
Endemic controls	23	28.5±13.1	7–47
Males	6	18.1±4.4	13–25
Females	17	32.2±13.2	7–47

<https://doi.org/10.1371/journal.pntd.0007444.t001>

containing 0.1% Triton-X-100 and kept at room temperature for at least 30 minutes for fixation. Samples were then filtered through two layers of gauge mesh, centrifuged 1500 g for 3 minutes, and the supernatant discarded. Seven ml of 10% formalin was added followed by 3 ml ethyl acetate, and samples shaken for 30 seconds and centrifuged 1500 g for 3 minutes. Supernatants were discarded and saline added to the sediment, which was then analysed under a light microscope at 10x magnification for the presence of helminth eggs.

DNA extraction

A portion (200µg) of each stool sample (as above) was transferred to OMNIgene GUT tubes (DNA Genotek) following the OMNIgene GUT protocol for transferring specimens from the collection tubes to this kit (protocol PD-PR-00434). The specimen was homogenised and stabilised within these tubes prior to DNA extraction. Genomic DNA was extracted with the QIAamp PowerFecal DNA Kit (QIAGEN) following the manufacturer's protocol with the exception that proteinase K was added after the bead beating step and the sample incubated at 65°C for 30 minutes. This was to ensure removal of histones from eukaryotic DNA which could inhibit downstream amplification. The remainder of the protocol was carried out as per manufacturer's instructions. Purified genomic DNA was eluted in 100µl of PCR-grade water. Genomic DNA samples were shipped to the Telethon Kids Institute in Perth, Australia, at ambient temperature for further analysis and stored at -20°C upon arrival.

Sequencing controls

As a positive control for the 16S rRNA and 18S rRNA sequencing we used the ZymoBIOMICS Microbial Community DNA Standard (Integrated Sciences, Chatwood, NSW, Australia) which comprised DNA for 8 bacterial (*Listeria monocytogenes*—12%, *Pseudomonas aeruginosa*—12%, *Bacillus subtilis*—12%, *Escherichia coli*—12%, *Salmonella enterica*—12%, *Lactobacillus fermentum*—12%, *Enterococcus faecalis*—12%, *Staphylococcus aureus*—12%) and 2 eukaryotic (*Saccharomyces cerevisiae*—2%, and *Cryptococcus neoformans*—2%) species. To ensure that amplification of 16S rRNA and 18S rRNA amplicons worked well in the presence of human DNA, we included replicates of the positive control with human DNA spiked in at 0%, 0.1%, 0.5%, 1%, and 10% of the total 20 ng reaction, as indicated. For the 18S rRNA positive control we also spiked in DNA from two protozoan parasites *Leishmania major* (1.2 ng/20 ng reaction) and *Toxoplasma gondii* (1.2 ng/20 ng reaction) to some mock controls as indicated. Negative controls included 4 unused OMNIgene GUT tubes processed for DNA in the Indian laboratory using the QIAamp PowerFecal DNA Kit (QIAGEN) kits as for preparation of experimental sample DNAs. These were included in the preparation of amplicons for sequencing.

Amplicon sequencing

Genomic DNA was quantified with the Qubit 2.0 dsDNA HS assay kit (Invitrogen) and diluted to 10 ng/µl for amplicon PCR. Aliquots for sequencing of 16S rRNA gene were sent to the Australian Genome Research Facility (AGRF) for amplification, Nextera XT dual-indexing, pooling and sequencing. The V3-V4 region of the gene was amplified with primers 341F (5'CCTAYGGGRBGCASCAG) and 806R (5'GGACTACNNGGGTATCTAAT) [27]. Aliquots for sequencing of the 18S rRNA gene were first amplified with the V9 region primers 1391f (5'TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTACACACCGCCCGTC) and EukBr (5'GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGATCCTTCTGCAGG TTCACCTAC) from the Earth Microbiome Project (EMP) standard protocol (<http://www.earthmicrobiome.org/protocols-and-standards/18s/>), with Illumina Nextera adapters attached (adapters in bold). The mammalian blocking primer (5'GCCCCGTCGCTACTACCGATTGG/

ideoxyI//ideoxyI//ideoxyI//ideoxyI//ideoxyI/TTAGTGAGGCCCT/3SpC3/) was also included to reduce amplification of human 18S rRNA. Samples were amplified in duplicate with the PCR mixture and conditions specified in the EMP protocol, and duplicates were pooled after amplification. These amplicons were then sent to AGRF for purification, dual-indexing, pooling and sequencing.

Sequence analysis

Demultiplexed sequence data was received from AGRF in FASTQ format and imported into the QIIME2 v2018.11 [28] version of the QIIME software [29] for analysis. Full documentation of all analysis code in this project can be found at <https://rachaellappan.github.io/VL-QIIME2-analysis/>. In brief, for both the 16S rRNA and 18S rRNA data, amplicon primers were first removed with the cutadapt plugin [30]. Reads were denoised, filtered and trimmed (where the median quality score fell below 30) with the DADA2 plugin [31]. Amplicon sequence variants (ASVs) were classified within QIIME2 using the SILVA v132 database [32], with a classifier trained on the amplified region [33]. These ASVs equate to classifying operational taxonomic units (OTUs) based on 100% sequence identity. This differs from the traditional method of clustering sequences and calling OTUs based (usually) on 97% sequence identity [34]. Note that we use the terminology of ASVs throughout. The majority taxonomy with 7 levels was used for both the 16S rRNA and 18S rRNA datasets, i.e. this taxonomy is taken to species level, but since species-level identification was not complete we elected to use genus-level classifications for most analyses. *De novo* phylogenetic trees used in downstream measures of diversity that included phylogenetic distances (cf. below) were created using \log_{10} read counts and the phylogeny align-to-tree-mafft-fastree (MAFFT multiple sequence alignment program) [35, 36] plugin in QIIME2. Stacked bar plots of percent relative abundance of different taxa across samples were generated in QIIME2. Aggregate percent abundance of taxa at the population level was determined by summing read counts per taxon across all study participants and calculating these as a proportion of the total read count. Prior to calculating downstream measures of diversity 18S rRNA data was also filtered to remove a small number of mammalian 18S rRNA reads, and also to remove bacterial 16S rRNA reads due to non-specificity of primers for eukaryotic 18S rRNA as described in the results section.

Measures of diversity

Alpha (within-sample) diversity was measured using Shannon's Diversity [37], Faith's Phylogenetic Diversity (PD) [38], and Pilou's Evenness [39] indices, calculated per sample within QIIME2 using rarefied counts (i.e. subsampled to the same sequencing depth across samples; rarefied to 21,383 for the 16S dataset, and 2,319 for the filtered 18S dataset). Boxplot figures for alpha diversity were created using GraphPad Prism v8, with between group statistical differences determined using the Mann-Whitney or ANOVA nonparametric tests in Prism. Beta (across-sample) diversity carried out using rarefied counts (as above) was measured using the Bray-Curtis dissimilarity statistic [40] based on compositional dissimilarity between samples taking abundance into account, unweighted UniFrac distances that measures phylogenetic distances between taxa, or weighted UniFrac distances [41] that measures phylogenetic distances and also takes relative abundance into account, as indicated. Principal coordinates (PCoA) plots for these indices were either generated within QIIME2 using the EMPEROR graphics tools [42], or in QIIME1 version 1.9.1 [29] using the `make_2d_plots.py` script. Analysis of composition of microbiomes (ANCOM) [43], previously reported [44] as a sensitive method (for samples >20) with good control of false discovery rate, was also used within QIIME2 to determine between group differences in taxon abundance. The use of log ratios within this test takes

account of differences in sequencing depth. Significance is reported as a W-value which provides a measure of how significant a specific taxon is relative to a number of other taxa used in the analysis.

Results

Prokaryotic and eukaryotic taxa identified across the study sample

A total of 12.5 million raw paired-end reads were generated across 46 faecal samples (mean±SD 271,509±166,138 reads/sample) for 16S rRNA amplicons, and a total of 1.6 million raw paired-end reads (mean±SD 34,112±6,852 reads/sample) for 18S rRNA amplicons. After read pre-processing of 16S rRNA sequence data there were 91,923±69,706 joined paired-end reads (minimum read depth 21,383; 33±6% of raw paired-end reads retained) per sample across VL and EC samples. Rarefaction plots (Part (a) in [S1 Fig](#)) based on Shannon's diversity index confirmed equivalent alpha diversity across the range of read depths from 2,000 to >20,000 (i.e. maximum alpha diversity is achieved at 2,000 reads). Using this read pre-processed data, 348 unique genera were identified from the 16S rRNA reads classified as ASVs within DADA2. The full taxonomy of these prokaryote 100% sequence identity ASVs is provided in [S2 Table](#). For 18S rRNA the initial pre-processed read depth across experimental samples ranged from 21,460–49,822 joined paired-end reads (95±2% of raw paired-end reads retained). However, preliminary ASV classification demonstrated high abundance of unclassified "Eukaryota" across all samples ([S2 Fig](#)). As can be seen from part (a) in [S2 Fig](#) VL cases and endemic controls were equally distributed across the range of abundances for this unclassified taxon. BLAST analysis of these unclassified taxa revealed alignment to bacterial 16S rRNA genes, demonstrating that the EMP primers for 18S rRNA amplification were not eukaryote-specific. These sequences were filtered out, leaving 77–27,315 joined paired-end reads per sample across the 46 samples. Using this filtered dataset, 62 unique genera were identified from the 18S rRNA reads classified as ASVs within DADA2. The full taxonomy of these eukaryotic 100% sequence identity ASVs is provided in [S3 Table](#). Shannon's index rarefaction plots (Part (b) in [S1 Fig](#)) indicated equivalent alpha diversity across read depths ≥400 (i.e. maximum alpha diversity that it was possible to determine in our samples was achieved at ≥400 reads). However, to provide robustness in downstream analyses of diversity a rarefied read depth of 2,319 was used. This resulted in loss of N = 8 VL and N = 7 EC samples for whom read depths were <1000 reads, leaving sample sizes of 15VL and 16EC in the downstream eukaryotic analyses.

[Fig 1\(a\)](#) provides a descriptive overview of bacterial gut microflora in our population in the form of a QIIME2-generated bar plot of 16S rRNA-identified bacterial taxa ordered by abundance across the 46 samples. [Fig 1\(b\)](#) provides the bar plot for the mock positive controls (see also Part (c) in [S2 Fig](#) for negative controls), demonstrating equivalent relative abundance of control bacterial taxa from the ZymoBIOMICS Microbial Community DNA Standard in the presence of increasing concentrations of human DNA. [Table 2A](#) shows the aggregated relative abundance at genus-level (% total counts and family provided in parentheses) for the 8 most prevalent taxa identified in the study sample. This includes *Prevotella* 9 (33.42%; Prevotellaceae), *Faecalibacterium* (11.32%; Ruminococcaceae), *Escherichia-Shigella* (9.14%; Enterobacteriaceae), *Alloprevotella* (4.46%; Prevotellaceae), *Bacteroides* (4.07%; Bacteroidaceae), *Prevotella* 2 (3.61%; Prevotellaceae), *Ruminococcaceae* UCG-002 (1.58%; Ruminococcaceae), and *Bifidobacterium* (1.54%; Bifidobacteriaceae). [Fig 1\(c\)](#) provides a heatmap of relative abundances for the top genera (*Prevotella*, *Faecalibacterium*, *Bacteroides* and *Escherichia-Shigella*) that relate to previously described enterotypes discussed in more detail in the discussion section below [[19](#), [45](#)]. Our study sample suggests a predominance of enterotype-2 driven by

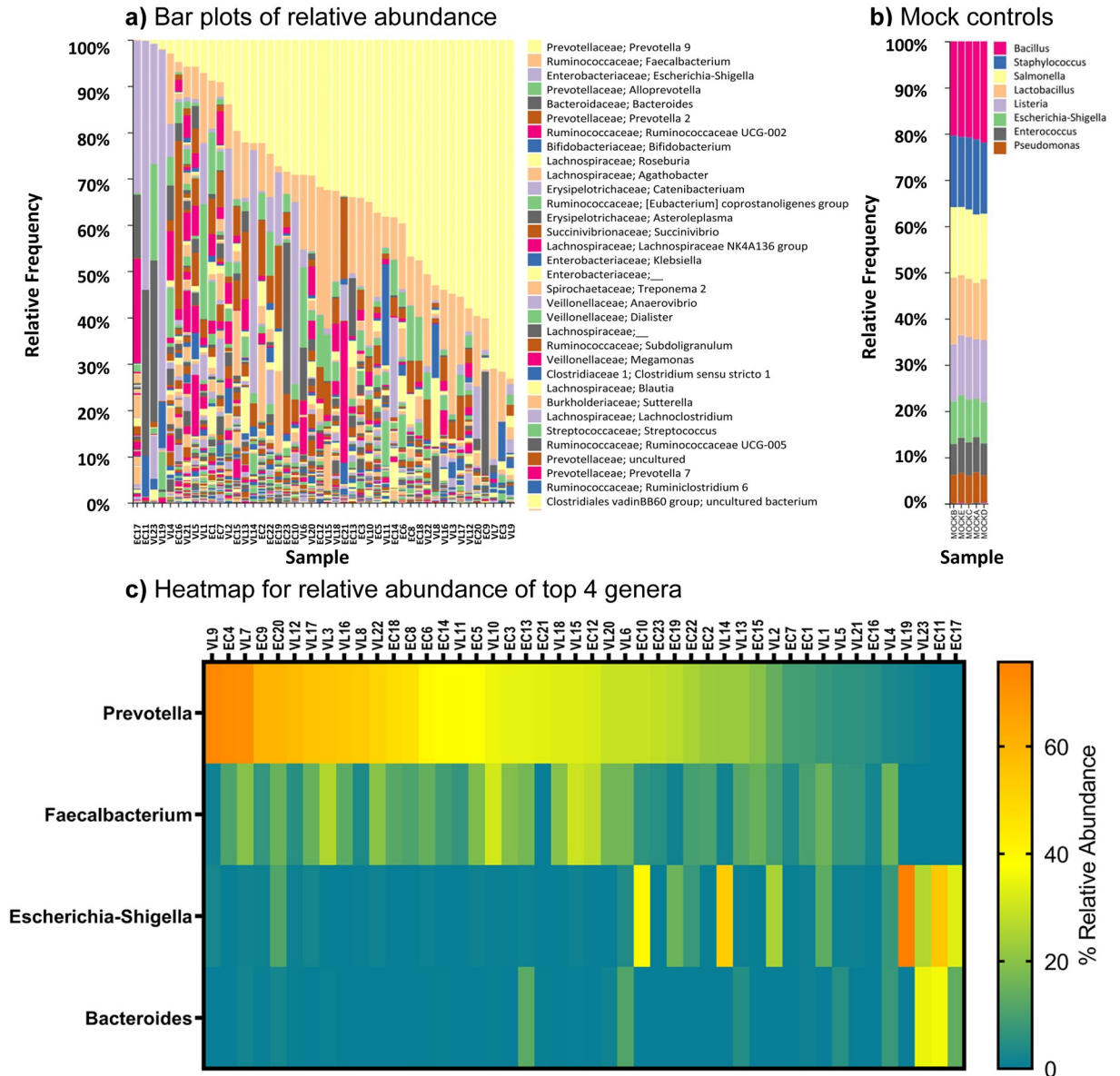


Fig 1. Results of 16S rRNA analysis of ASVs at genus level in the study sample. (a) Bar plot ordered by relative abundance for prokaryotic taxa colour coded according to the key to the right of the bar plot. (b) Bar plot for prokaryotic mock control samples. (c) Heat map of relative abundance for the 4 taxa associated with previously defined bacterial enterotypes, ordered by relative abundance of *Prevotella*.

<https://doi.org/10.1371/journal.pntd.0007444.g001>

Prevotella, with only a small number of individuals conforming to enterotype-1 driven by *Bacteroides* and a lack of enterotype-3 driven by *Ruminococcus*. High *Faecalibacterium* abundance was generally associated with high *Prevotella*, whereas *Escherichia-Shigella* was associated with low *Prevotella*. Bacterial taxa (e.g. *Dialister*, *Megasphaera*, *Mitsuokella*, *Lactobacillus*) previously shown to be characteristic of Indian gut microbiomes [20, 24, 46] occurred at <1% in our study sample (Table 2).

Fig 2(a) provides a bar plot of 18S rRNA-identified eukaryotic taxa ordered by abundance across the 46 samples. Fig 2(b) provides the bar plot for the mock positive controls, demonstrating equivalent relative abundance of the two control fungal eukaryotes in the ZymoBIOMICS

Table 2. Genus-level community composition, and family to which they belong, for VL cases, endemic controls, and the total study sample. Abundance at the study group level was calculated for aggregated read counts across all samples for ASVs classified at genus level (i.e. aggregates all taxa with the same genus assignment). Genera with an average group-level abundance <1% are not shown, except where relevant to the results or discussion. Full details are provided in S2 and S3 Tables.

Genus level taxonomy	Family level taxonomy	Percent in VL cases	Percent in endemic controls	Percent across all samples
A. Prokaryotes based on 16S rRNA sequencing				
<i>Prevotella</i> (2+7+9)	Prevotellaceae	40.77	33.61	37.47
<i>Faecalibacterium</i>	Ruminococcaceae	12.44	10.02	11.32
<i>Escherichia-Shigella</i>	Enterobacteriaceae	8.82	9.51	9.14
<i>Alloprevotella</i>	Prevotellaceae	4.31	4.63	4.46
<i>Bacteroides</i>	Bacteroidaceae	3.46	4.78	4.07
<i>Ruminococcaceae UCG-002</i>	Ruminococcaceae	1.87	1.24	1.58
<i>Bifidobacterium</i>	Bifidobacteriaceae	1.67	1.38	1.54
<i>Roseburia</i>	Lachnospiraceae	0.68	2.38	1.47
<i>Agathobacter</i>	Lachnospiraceae	0.47	2.43	1.38
<i>Catenibacterium</i>	Erysipelotrichaceae	0.73	1.76	1.21
<i>Asteroleplasma</i>	Erysipelotrichaceae	0.02	2.45	1.14
<i>Succinivibrio</i>	Succinivibrionaceae	0.71	1.38	1.02
<i>Clostridiales vadin BB60 group</i>	Order: Clostridiales	1.38	0.31	0.89
<i>Anaerovibrio</i>	Veillonellaceae	0.01	1.48	0.69
<i>Dialister</i>	Veillonellaceae	0.41	0.91	0.64
<i>Megamonas</i>	Veillonellaceae	0.25	0.99	0.59
<i>Megasphaera</i>	Veillonellaceae	0.57	0.14	0.37
<i>Mitsuokella</i>	Veillonellaceae	0.07	0.13	0.10
<i>Lactobacillus</i>	Lactobacillaceae	0.10	0.08	0.09
<i>Ruminococcaceae UCG-014</i>	Ruminococcaceae	0.05	0.28	0.15
<i>Gastranaerophilales_ uncultured bacterium</i>	Order: Gastranaerophilales	0.002	0.24	0.15
B. Eukaryotes based on 18S rRNA sequencing				
<i>Blastocystis</i> sp. MJ99-568	Order: Stramenopiles	38.88	25.48	32.27
<i>Blastocystis</i> ambiguous taxon	Order: Stramenopiles	21.32	29.78	25.49
<i>Dientamoeba</i>	Tritrichomonadea	9.38	14.94	12.12
<i>Pentatrichomonas</i>	Trichomonodea	18.76	1.29	10.13
<i>Entamoeba</i>	Entamoebida	0.92	6.04	3.45
Ascaridida	Chromodorea	0	1.58	0.78
Rhabditida	Chromodorea	0.86	0.69	0.77
Cyclophyllidea	Eucestoda	0	0.38	0.19

<https://doi.org/10.1371/journal.pntd.0007444.t002>

Microbial Community DNA Standard, together with identification of *Leishmania* and *Toxoplasma* DNA spiked into some control samples in-house. Full details of positive and negative control samples are provided in S2 Fig. Table 2B shows the aggregated relative abundance at genus or lowest order level classified (higher order classification also provided in parentheses) for the eukaryotic taxa of interest (i.e. excluding plant genera) identified in the study sample. This includes *Blastocystis* sp. MJ99-568 (32.37%; Order: Stramenopiles), *Blastocystis* ambiguous taxon (25.49%; Order: Stramenopiles), *Dientamoeba* (12.12%; Family: Tritrichomonadea), *Pentatrichomonas* (10.13%; Family: Trichomonodea), *Entamoeba* (3.45%; Family: Entamoebida), Ascaridida (0.78%; Family: Chromodorea; likely *Ascaris* round worm cf. below), Rhabditida (0.77%; Family: Chromodorea; likely *Strongyloides* or hookworm cf. below), and Cyclophyllidea (0.19%; Order: Eucestoda; likely *Hymenolepis* tapeworm cf. below). Of these, unsupervised cluster analysis performed using the QIIME2 feature-table heatmap plugin in which both axes are clustered with the average linkage method using the Bray-Curtis distance metric (Fig 2(c))

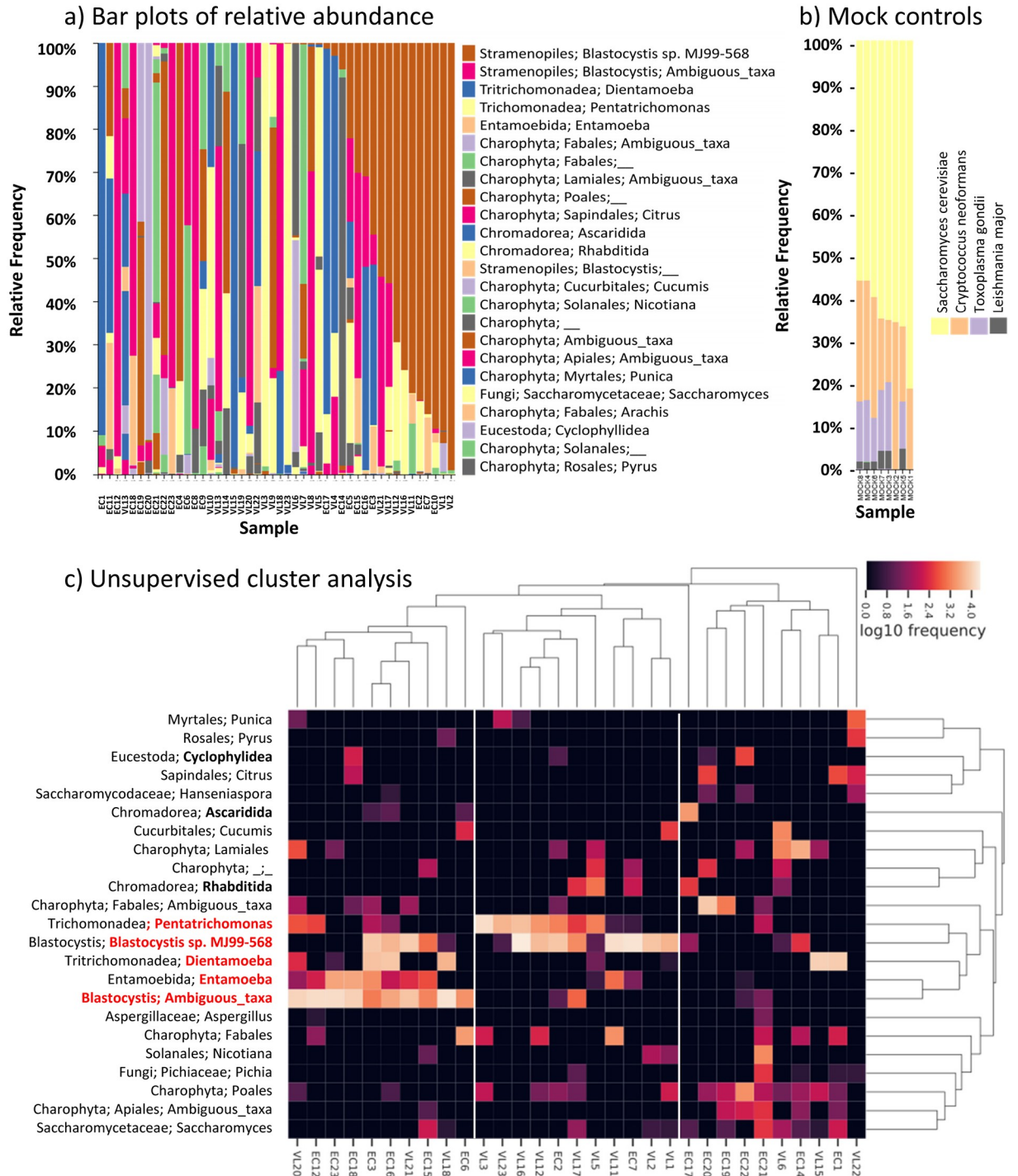


Fig 2. Results of 18S rRNA analysis of ASVs at genus level in the study sample. (a) Bar plot ordered by relative abundance for eukaryotic taxa colour coded according to the key to the right of the bar plot. (b) Shows the bar plot for eukaryotic mock control samples; mock controls 1 and 2 were not spiked with *Leishmania* or *Toxoplasma* DNA. (c) Heatmap derived from unsupervised cluster analysis of eukaryotic taxa. Red labelling indicates the cluster of 5 most prevalent eukaryotic protozoan taxa for which samples cluster into three eukaryotic enterotypes according to: (A) *Blastocystis* Ambiguous taxon with/without *Dientamoeba* and/or *Entamoeba*; (B) *Pentatrachomonas* and/or *Blastocystis* sp. MJ99-568 in the absence of taxa from cluster A; (C) low abundance of these protozoan taxa. Bold labelling indicates the helminth eukaryotic taxa. Other eukaryotic taxa include fungi and a variety of plant taxa derived from the diet of participants. Of note, only one individual was observed with *Leishmania* DNA in the faeces (which does not show up in the bar plot of abundance for the top 24 taxa).

<https://doi.org/10.1371/journal.pntd.0007444.g002>

shows 3 clusters of individuals that are defined by high relative frequencies of: (A) N = 10 (3 VL; 7 EC) *Blastocystis* Ambiguous taxon with/without *Dientamoeba* and/or *Entamoeba*; (B) N = 11 (9 VL; 2 EC) with *Pentatrichomonas* and/or *Blastocystis* sp. MJ99-568 in the absence of taxa from cluster A; and (C) N = 10 (3VL; 7EC) absence or low abundance of all 5 of these taxa but high frequency of *Saccharomyces* and plant species. For easy reference, we tentatively refer to these below as “eukaryotic enterotypes” A, B and C with the caveat that further work is required to further define eukaryotic enterotypes. Species level identification carried out as part of the overall QIIME2 taxonomic analysis confirmed *Pentatrichomonas* as *P. hominis* and *Dientamoeba* as *D. fragilis*. *Entamoeba* was not identifiable to species level.

Influence of age, sex or microgeographic location on gut microflora in the study sample

Other studies in India [21, 23] and elsewhere [18, 45] have demonstrated changes in the composition of gut flora with age. In our study it was important to determine whether there were any global effects of age and/or gender on microbial diversity that needed to be adjusted for in analysis of differences between VL cases and EC. Parts (a-b) in S3 Fig provides bar plots of the 16S rRNA data reordered by age or gender (and secondarily on relative abundances of taxa; key as per Fig 1(a)), respectively. Visual inspection suggests no overt effects of either age or sex on composition of microbial flora. This is confirmed comparing species richness (Number of ASVs) and alpha diversity measures (Faith’s PD, Pielou Evenness, Shannon’s diversity index) between genders (Parts (c-f) in S3 Fig) and beta diversity PCoA plots for weighted UniFrac distances by gender (Part (g) in S3 Fig) and for age as a continuous variable (Part (h) in S3 Fig). Note all comparisons have non-significant adjusted p-values. Similarly, parts (a-b) in S4 Fig provide bar plots of the 18S rRNA data reordered by age and gender, respectively. Again, visual inspection suggests no overt effects of either age or sex on composition of microbial flora. This is supported by comparisons of species richness (Number of ASVs) and alpha diversity measures (Faith’s PD, Pielou Evenness, Shannon’s diversity index) between sexes (Parts (c-f) in S4 Fig) and PCoA plots for weighted UniFrac distances by gender (Part (g) in S4 Fig) and for age as a continuous variable (Part (h) in S4 Fig). In this case there was borderline significance ($p = 0.048$) for difference between genders for Faith’s phylogenetic diversity (all other p-values were not significant). On balance, we conclude there are no overt independent effects of age or sex on either 16S rRNA-determined bacterial diversity or 18S rRNA-determined eukaryotic diversity in gut flora that would affect analysis of microbial profiles comparing VL cases with EC.

Substructure in microbial diversity due to microgeographic location could also confound downstream analysis comparing VL cases and EC. Generally, EC were matched to VL for location as they were attendants of patients at KAMRC. KAMRC is a centre that draws patients from across Bihar State in India, but principally (in this study) from within the district of Muzzaffapur. Part (a) in S5 Fig provides a map of subdistricts or blocks within Muzzaffapur District showing participant locations. Part (b) in S5 Fig shows the 16S rRNA relative abundance bar plot reordered by subdistrict (key to taxa as per Fig 1(a)). Box plots (Parts (c-f) in S5 Fig) comparing species richness (Number of ASVs) and alpha diversity measures (Faith’s PD, Pielou Evenness, Shannon’s diversity index) across blocks where $N \geq 4$ indicate no statistically significant differences by subdistrict (Anova not significant). Principle coordinate plots based on weighted UniFrac distances (Part (g) in S5 Fig) also show no overt clustering by subdistrict. Similar results were obtained for the 18S rRNA data (S6 Fig). We conclude there are no overt differences in gut microflora by subdistrict in this region of Bihar State.

Comparing gut microflora between VL cases and EC in the study sample

Fig 3(a) and 3(b) provide bar plots of the relative abundances of taxa reordered by VL case or EC status for 16S rRNA data and 18S rRNA data, respectively. Species richness (Number of ASVs) and alpha diversity box plots (Faith's PD, Pielou Evenness, Shannon's diversity index) comparing VL cases with EC (Fig 3(c)–3(f) for 16S rRNA; Fig 3(g)–3(j) for 18S rRNA) show no statistically significant differences in alpha diversity between these two groups. Beta-diversity PCoA plots based on Bray-Curtis distances (Fig 4(a) and 4(b)) that take relative abundance into account or using weighted UniFrac distances (Fig 4(c) and 4(d)) that take abundance and phylogenetic distances between taxa into count, also show no clear evidence for clustering based on VL cases *versus* EC status.

Although there were no global differences in gut prokaryotic or eukaryotic microbial composition between VL cases and EC as determined by alpha and beta diversity measures, analysis of 16S rRNA data using ANCOM identified two bacterial genera, Ruminococcaceae UCG-014 (Family: Ruminococcaceae) and Gastranaerophilales_uncultured bacterium (Phylum: Cyanobacteria; Class: Melainabacteria; Order Gastranaerophilales), that were enriched in EC compared to VL cases (Fig 5(a) and 5(b); W-values of 61 and 41, respectively). These differences are also apparent in the aggregated relative abundance data presented in Table 2, with Ruminococcaceae UCG-014 at 0.28% in EC compared to 0.05% in VL cases and Gastranaerophilales_uncultured at 0.24% in EC compared to 0.002% in VL cases.

Analysis of 18S rRNA genus-level data using ANCOM did not show differences between VL cases compared to EC. However, visual inspection of the 18S rRNA bar plots (Fig 2(a)) suggested a trend for higher abundance of *P. hominis* in VL cases compared to EC. In contrast, there were more individuals with higher relative abundance of *Entamoeba* in EC than in VL cases. This is also reflected in the aggregated relative abundance data for eukaryotic genera presented in Table 2B, and further highlighted on a heatmap (Fig 6(a)) for eukaryotic protozoa \log_{10} read counts ordered by *P. hominis* counts in EC and VL cases. This aligns with the "eukaryotic enterotypes" defined above, where 7/9 individuals classified as "eukaryotic enterotype B" defined by relative frequencies of read counts for *P. hominis* and *Blastocystis* sp. MJ99-568 were VL patients, and conversely 7/10 individuals classified as "eukaryotic enterotype A" defined by *Blastocystis* ambiguous taxon, with/without *Entamoeba* or *D. fragilis*, were EC. Comparisons of \log_{10} read counts (Fig 6(b) and 6(c); $p = 0.05$) and relative abundance (Fig 6(d) and 6(e); $p = 0.01$) show that VL cases positive for *P. hominis* had significantly higher *P. hominis* burdens compared to EC, whereas this was not the case for the between group comparison of individuals positive for *Entamoeba* (p -values not significant) This suggests that when VL patients acquire *P. hominis* they have a more intense colonisation whereas when they acquire *Entamoeba* the relative abundance is the same as for EC individuals.

Influence of known enteric pathogens on microbial diversity in gut flora

Entero-pathogens have been observed to perturb bacterial diversity in gut flora, for example in the context of a bacterial enterotype defined by *Escherichia-Shigella* abundance [45]. We therefore looked across our sample to see if such effects could be observed in our study population. Looking across relative abundance plots for 16S rRNA data (see Figs 1 and 3) for VL cases and EC there were no overt group-specific differences in relative abundance of bacterial taxa that could contain pathogenic species, e.g. *Escherichia-Shigella*. However, we did observe that some individuals positive for *Escherichia-Shigella* appeared to have reduced diversity in terms of other bacterial species in their microflora.

Another eukaryotic genus that contains potentially pathogenic species is *Blastocystis*. In looking for effects on the bacterial microflora of eukaryotic taxa that might contain putative

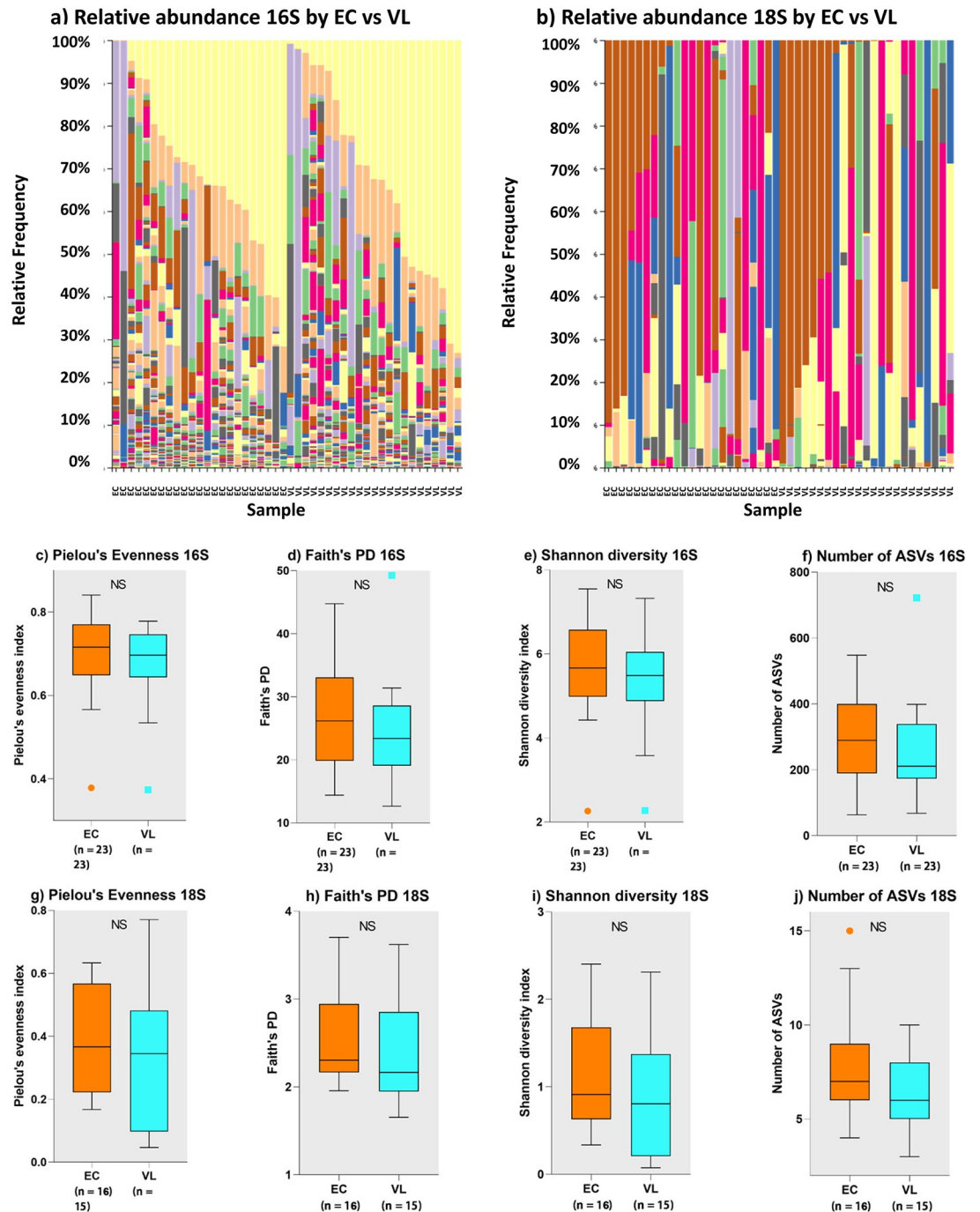


Fig 3. Analysis of gut microbial profiles in relation to VL case versus EC status. On 16S rRNA-determined prokaryotic and 18S rRNA-determined eukaryotic gut microbial profiles. (a) bar plots for relative abundance of 16S rRNA taxa separated for VL cases and EC groups, each ordered by relative abundance of taxa; for colour coding see key to main Fig 1(a). (b) bar plots for relative abundance of 18S rRNA taxa separated for VL cases and EC groups, each ordered by relative abundance of taxa; for colour coding see key to main Fig 2(a). (c) to (f) comparisons of species richness (number of ASVs) and alpha diversity measures (as labelled) for 16S rRNA data. (g) to (j) comparisons of species richness (number of ASVs) and alpha diversity measures (as labelled) for 18S rRNA data.

<https://doi.org/10.1371/journal.pntd.0007444.g003>

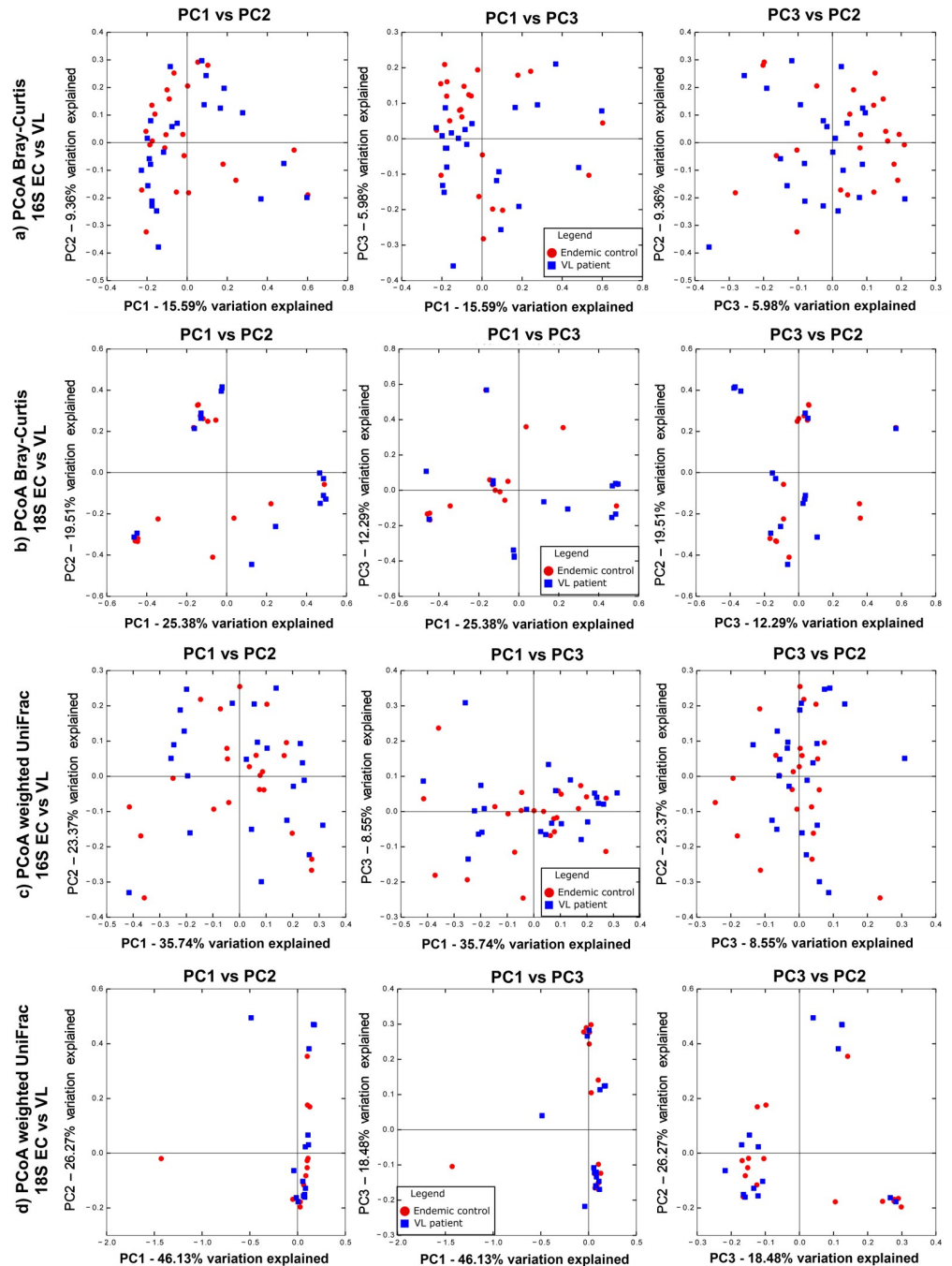


Fig 4. Beta diversity comparisons for VL cases and EC. (a) and (b) are PCoA plots (left to right: PC1xPC2; PC1xPC3; PC2xPC3) for Bray-Curtis dissimilarity measures for 16S rRNA and 18S rRNA data, respectively. (c) and (d) are PCoA plots (left to right: PC1xPC2; PC1xPC3; PC2xPC3) for weighted UniFrac measures for 16S rRNA and 18S rRNA data, respectively. See keys on plots for colour coding of VL cases and EC. Note: the weighted UniFrac distances are skewed due to 2 samples with high worm burdens.

<https://doi.org/10.1371/journal.pntd.0007444.g004>

pathogens, we noted clustering according to *Blastocystis* abundance on Jaccard index beta diversity PCoA plots (Fig 7(a)–7(c)) and unweighted Unifrac PCoA plots (Fig 7(d)). This was true for both *Blastocystis* sp. MJ99-568 (Fig 7(a)) and *Blastocystis_ambiguous* taxon (Fig 7(b)). Specifically, individuals with high *Blastocystis* abundance (46–98%) were almost exclusively

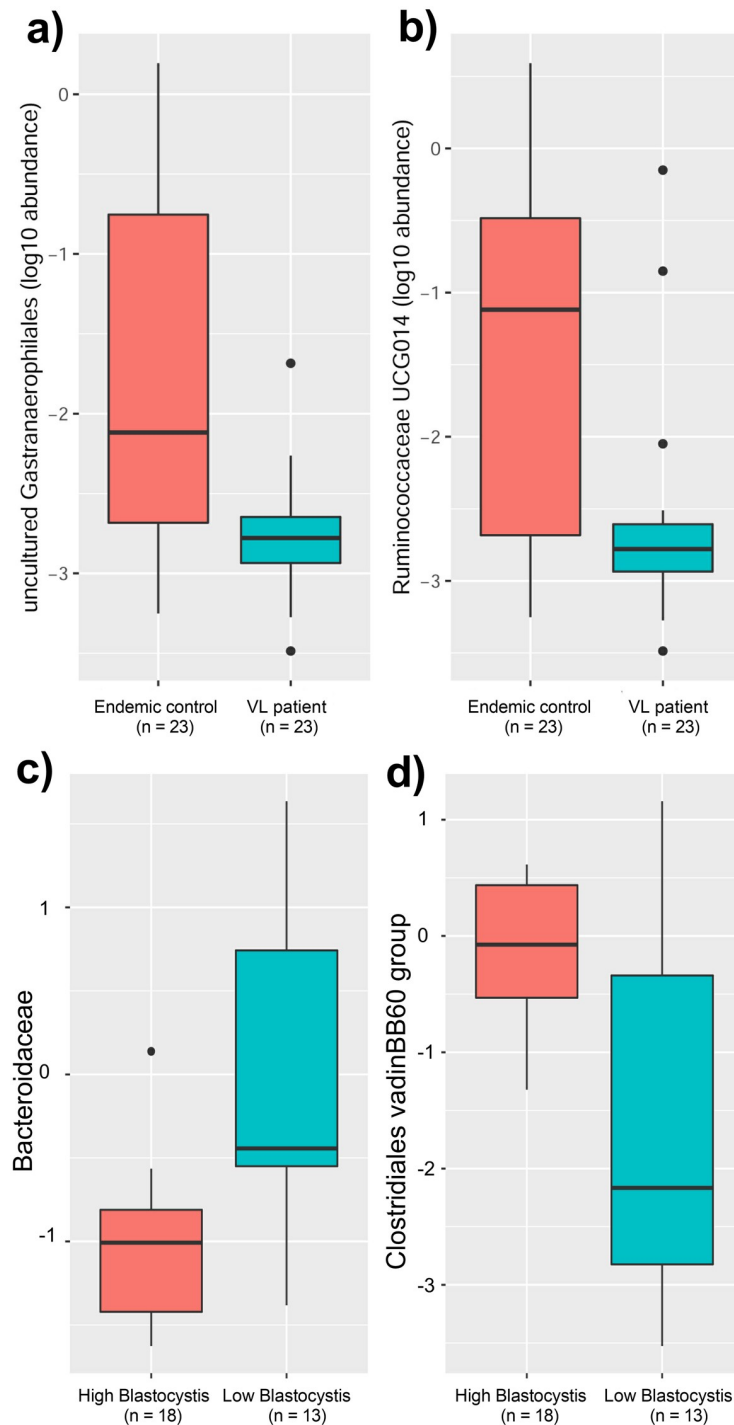


Fig 5. Results of ANCOM analysis of between-group differences in microbial composition. (a) and (b) box plots for prokaryotic taxa that differed between EC and VL cases. W-values were 41 for Gastranaerophilales presented in (a) and 61 for Ruminococcaceae presented in (b), indicating that these taxa were significantly higher in the EC group compared to VL in comparison with 41 or 61 other taxa examined, respectively. (c) and (d) ANCOM box plots for prokaryotic taxa that differed between groups of individuals carrying high (46–89%) versus low (<6%) relative abundance of eukaryotic Blastocystis. W-value = 31 (c) indicating that Bacteroides abundance was significantly lower in the high Blastocystis group in comparison with 31 other bacterial families examined. W-value = 30 (d) indicating that *Clostridiales vadinBB60* abundance was significantly higher in the high Blastocystis group in comparison with 30 other bacterial families examined.

<https://doi.org/10.1371/journal.pntd.0007444.g005>

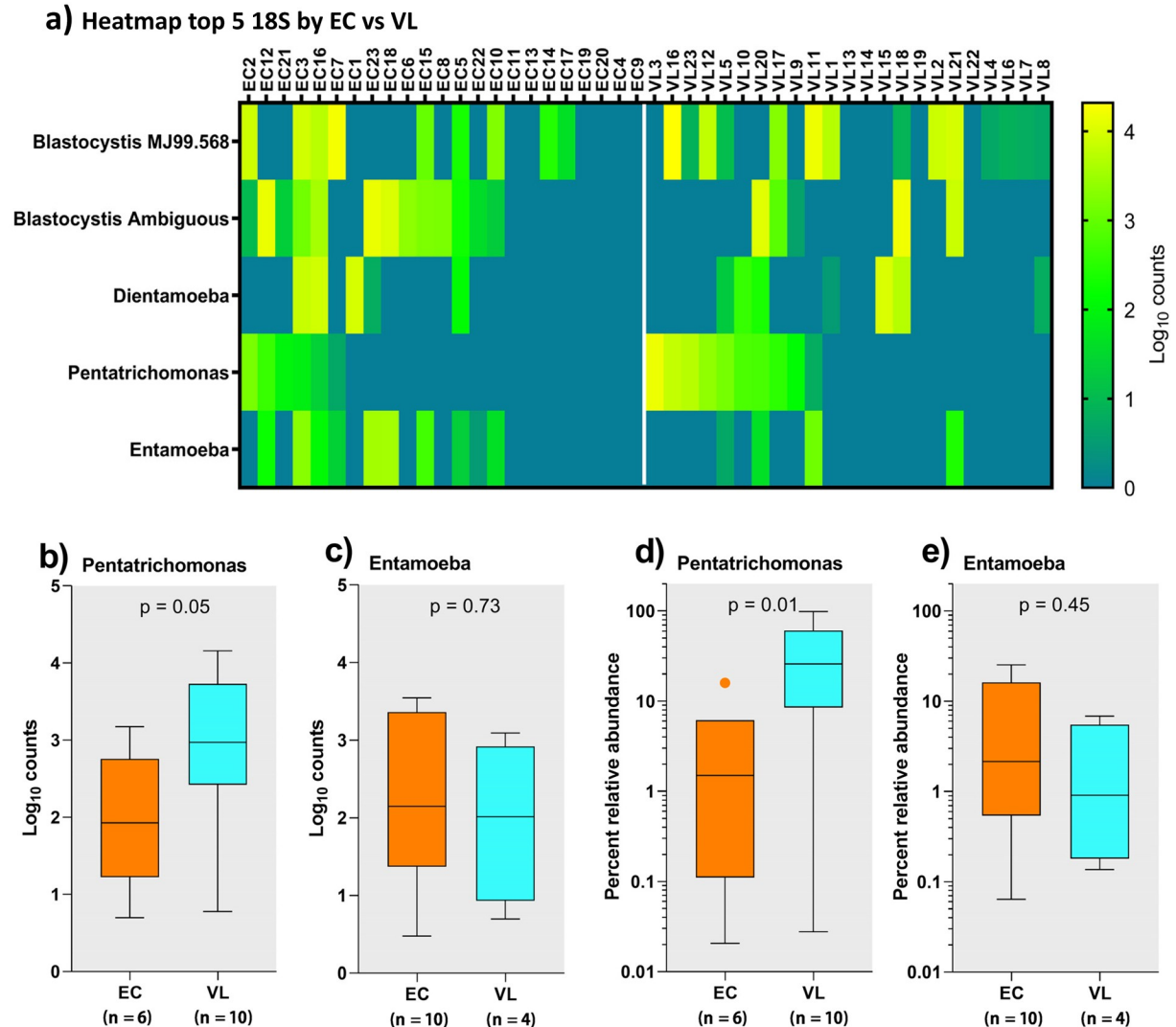


Fig 6. Influence of VL on eukaryotic *Pentatrachomonas* burden. (a) heatmap of the top 5 eukaryotic taxa separated by VL case versus EC status, each ordered according to *Pentatrachomonas* load. (b) and (c) box plots comparing *Pentatrachomonas* and *Entamoeba* log₁₀ counts between EC and VL cases. (d) and (e) show the same comparisons for percent relative abundance. VL cases carry higher loads of *Pentatrachomonas* but not of *Entamoeba*.

<https://doi.org/10.1371/journal.pntd.0007444.g006>

clustered to the right of the zero co-ordinate on PCoA axis 1, whereas individuals with low *Blastocystis* abundance (<6%) were clustered to the left. To determine what bacterial taxa might be driving this variation we carried out biplot analyses, results of which are superimposed onto PCoA plots of Jaccard indices based on presence/absence of bacterial taxa (Fig 7(a)–7(c)) and onto PCoA plots of unweighted UniFrac distances that take phylogenetic distances into account (Fig 7(d)). The four most significant bacterial taxa driven by eukaryotic *Blastocystis* abundance were *Escherichia-Shigella*, *Prevotella* 9, *Succinivibrio*, and *Asteroleplasma* for the Jaccard index analysis, and *Escherichia-Shigella*, *Bacteroides thetaiotaomicron*, *Faecalibacterium*, *Prevotella* 9 based on unweighted UniFrac distances. In both cases *Escherichia-Shigella* had the strongest effect (as indicated by the length of arrow on the biplots), which manifests as a negative correlation (slope = -0.96; $r^2 = 0.18$; $p = 0.003$) between log₁₀ counts for *Escherichia-Shigella* and *Blastocystis* (Fig 7(e)). We therefore conclude that, as previously

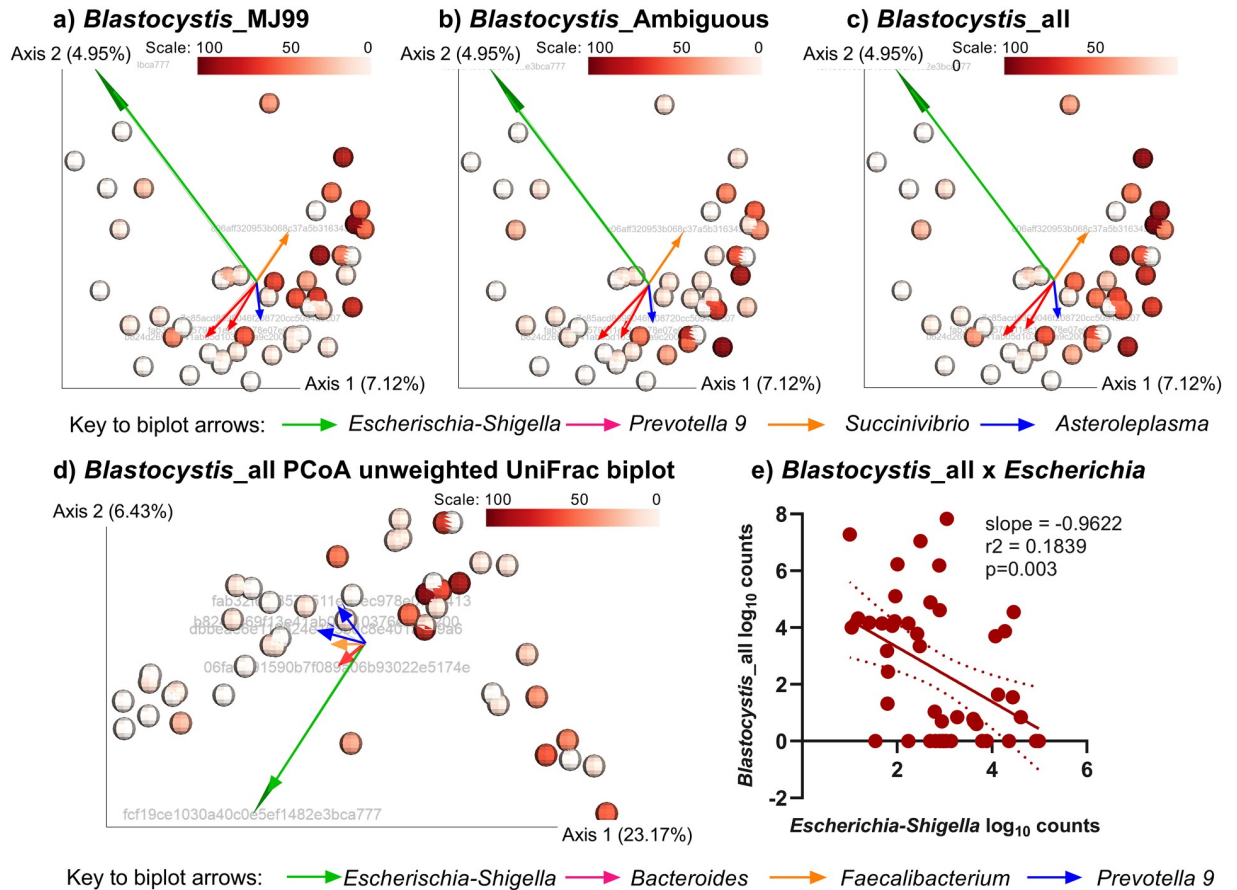


Fig 7. Influence of *Blastocystis* on beta diversity of bacterial taxa. (a-c) PCoA plots of Jaccard indices for beta diversity of bacterial microflora colour coded (see scales) according to relative abundances for (a) *Blastocystis_MJ99-340*, (b) *Blastocystis_Ambiguous* taxon, and (c) all *Blastocystis*. Results of the biplot analysis are superimposed on the PCoA plots, with length of arrows indicating the degree of influence of different bacterial taxa (see key to biplot arrows) on *Blastocystis* abundance. (d) PCoA plot of unweighted UniFrac distances colour coded (see scale) according to *Blastocystis* load, with biplot vectors superimposed as before. (e) graph showing negative correlation between *Blastocystis* log₁₀ counts and *Escherichia-Shigella* log₁₀ counts.

<https://doi.org/10.1371/journal.pntd.0007444.g007>

observed [45], high abundance of *Escherichia-Shigella* may be influencing dysbiosis of gut bacterial microflora in this population.

Blastocystis relative abundance also had a more general influence on alpha-diversity measures (Fig 8(a)–8(c) for *Blastocystis* MJ99-568, p-values <0.001 and 0.004 for Faith’s PD and Shannon diversity index, respectively; Fig 8(e)–8(g) for *Blastocystis* Ambiguous taxon, p-values 0.007, 0.013 and 0.003 for Pielou’s evenness index, Faith’s PD, and Shannon diversity index, respectively) and on species richness (Number of ASVs) (Fig 8(d) for *Blastocystis* MJ99-568, p<0.001; Fig 8(h) for *Blastocystis* Ambiguous taxon, p = 0.008). Notably, high abundance of both *Blastocystis* taxa was associated with higher species richness and higher measures of alpha diversity. To further identify specific bacterial taxa that might be related to the effects of *Blastocystis* on bacterial diversity we employed ANCOM to look at the family-level for differential 16S rRNA-determined bacterial abundance between high versus low *Blastocystis* carriers. This identified two bacterial taxa associated with high versus low *Blastocystis* count (Fig 5(c) and 5(d)). Bacteroidaceae (aggregate abundance 4.07%, see Table 2) were less abundant with high *Blastocystis* and more abundant with low *Blastocystis* (Fig 5(c); W-value = 31). Conversely,

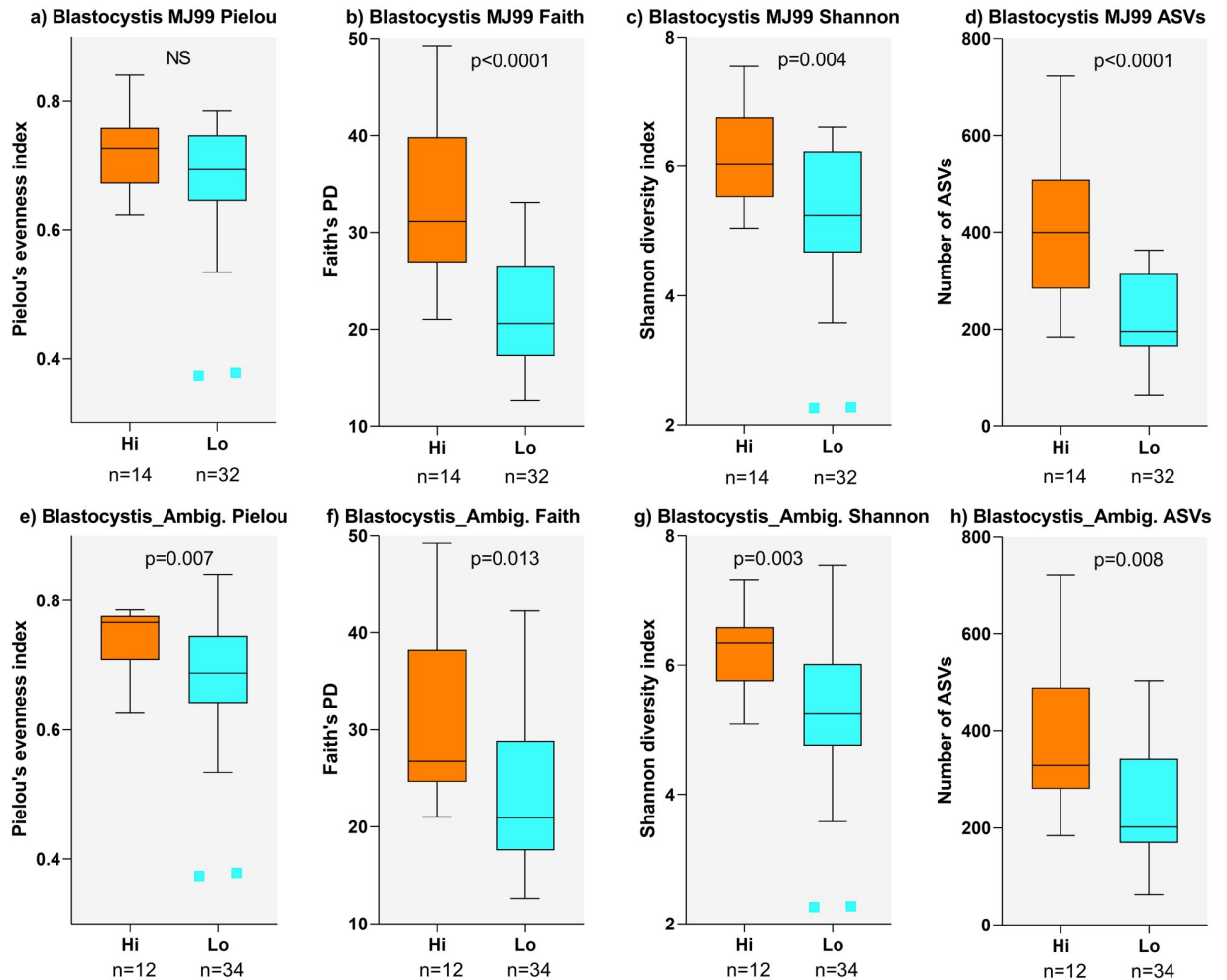


Fig 8. Influence of *Blastocystis* on alpha diversity of bacterial taxa. (a-d) box plots comparing influence of high or low *Blastocystis* MJ99-340 abundance on alpha diversity measures (as labelled) and species richness (as measured by number of ASVs). (e-h) the same comparisons for groups of individuals with high or low *Blastocystis*_Ambiguous taxon abundance.

<https://doi.org/10.1371/journal.pntd.0007444.g008>

Clostridiales vadinBB60 group (aggregate abundance 0.89%, see Table 2) were more abundant with high *Blastocystis* and less abundant with low *Blastocystis* (Fig 5(d); W-value = 30).

Using 18S rRNA analysis to identify gut helminths

One important reason for carrying out analysis of taxa based on 18S rRNA sequencing was to determine if it would be sufficiently sensitive to detect helminth species in faecal samples. Notwithstanding loss of read depth and sample numbers due to bacterial contamination of reads, we were able to detect 4 taxa containing pathogenic helminths, namely Ascaridida (*Ascaris*), Trichocephalida (*Trichuris*), Rhabditida (Hookworm–*Ancylostoma*) and Cyclophyllidea (*Hymenolepis*). Table 3 compares data based on microscopy and 18S rRNA analysis for those individuals (N = 36) for whom both datasets were available. Analysis using 18S rRNA identified 3/3 positive for *Ascaris* by microscopy, 0/8 positive for *Trichuris* by microscopy (with one that was positive for 18S rRNA but negative for microscopy), 7/7 positive for hookworm by microscopy (plus one positive by 18S rRNA and negative by microscopy), and 4/4 positive for *Hymenolepis* by microscopy.

Table 3. Comparison of helminth identification by microscopy (yes/no for presence/absence) with log₁₀ read counts for parallel identification to the appropriate order using 18S rRNA analysis. Bold indicates where positive microscopy and 18S rRNA concur; italics indicates where they do not concur.

Sample number	Ascaris	Ascaridida	Trichuris trichura	Trichocephalida	Hookworm	Rhabditida	Hymenolepis	Cyclophyllidea
EC01	No	0	No	0	No	0	No	0
EC02	No	0	No	0	No	0	Yes	0.90
EC03	Yes	0.78	No	0	No	0	No	0
EC04	No	0	No	0	Yes	2.44	No	0
EC05	No	0	Yes	0	Yes	2.43	No	0
EC06	Yes	1	Yes	0	No	0	No	0
EC08	No	0	Yes	0	No	0	No	0
EC09	No	0	No	0	No	0	No	0
EC10	No	0	No	0	Yes	2.06	No	0
EC11	No	0	No	0	No	0	No	0
EC12	No	0	No	0	No	0	No	0
EC14	No	0	No	0	No	0	No	0
EC15	No	0	No	0	No	0	No	0
EC17	Yes	3.41	No	0	Yes	2.52	No	0
EC18	No	0	No	0	No	0	Yes	2.24
EC19	No	0	No	0	No	0	No	0
EC20	No	0	No	0	No	0	Yes	0.78
EC21	No	0	No	0	No	0	No	0
EC22	No	0	No	0	No	0	Yes	2.64
EC23	No	0	No	0	No	0	No	0
VL01	No	0	No	0	No	0	No	0
VL02	No	0	No	0	No	0	No	0
VL03	No	0	No	0	No	0	No	0
VL04	No	0	No	0	No	0	No	0
VL05	No	0	Yes	0	Yes	3.04	No	0
VL06	No	0	No	0	Yes	1.41	No	0
VL08	No	0	No	0	No	0	No	0
VL09	No	0	Yes	0	No	0	No	0
VL11	No	0	Yes	0	No	0	No	0
VL12	No	0	No	<i>0.60</i>	No	0	No	0
VL14	No	0	No	0	Yes	1.34	No	0
VL16	No	0	Yes	0	No	0	No	0
VL17	No	0	Yes	0	No	2.42	No	0
VL20	No	0	No	0	No	0	No	0
VL21	No	0	No	0	No	0	No	0
VL22	No	0	No	0	No	0	No	0

<https://doi.org/10.1371/journal.pntd.0007444.t003>

Discussion

In this study we have used a meta-taxonomic approach to determine the composition of prokaryotic and eukaryotic gut microflora in a region of Bihar State in North-East India endemic for VL. Amplicon-based sequencing of the V3-V4 region of the 16S rRNA gene showed that the most abundant bacterial taxa identified in faecal samples from this region of India were *Prevotella* (37.1%), *Faecalibacterium* (11.3%), *Escherichia-Shigella* (9.1%), *Alloprevotella* (4.5%), *Bacteroides* (4.1%), *Ruminococcaceae* UCG-002 (1.6%), and *Bifidobacterium* (1.5%). Our study was also unique in using alternative primers for amplicon-based sequencing of the V9 region of the 18s rRNA gene to identify eukaryotic microflora. Although we encountered

one issue with cross-amplification of 16S rRNA sequences due to sequence similarity of primers, we were successful in (a) blocking amplification of human 18S rRNA, and (b) in amplifying protozoan and metazoan sequences of pathogenic interest. In so doing we determined that the most prevalent protozoan taxa present in this region of India were *Blastocystis* sp. MJ99-568 (32.37%), *Blastocystis* ambiguous taxon (25.49%), *D. fragilis* (12.12%), *P. hominis* (10.13%), and *Entamoeba* (3.45%).

Others have also used sequencing of 16S rRNA amplicon libraries to study the bacterial gut microbiome of people living in India [20–24, 46]. A study [20] of the gut microbiomes of 43 healthy individuals from urban and semi-urban districts in Maharashtra state (in the West of India, including Mumbai) also found that the microbial community at genus level was dominated by *Prevotella* (34.7%), *Bacteroides* (15.2%), and *Faecalibacterium* (5.6%), with additional prominent species *Megasphaera* (4.7%) and *Dialister* (3.9%). Similarly, the genera *Prevotella* (4.5%) and *Megasphaera* (8.5%) predominated in the gut microflora of 34 healthy Indian subjects from two urban cities (Delhi from the North; Pune from the West, also near Mumbai) [46] and were reported to be a distinctive feature of Indian gut microbiota compared with western cultures and to be a component of sharing OTUs with omnivorous mammals [46]. Elsewhere in India [24], the gut microbiome of a cohort (N = 110) of healthy subjects from North-Central India (Bhopal City, Madhya Pradesh) consuming a primarily plant-based diet was dominated by *Prevotella*, whereas that of omnivorous individuals from Southern India (Kerala) showed more prominent associations with *Bacteroides*, *Ruminococcus* and *Faecalibacterium*. This aligned with previously described human gut microbiome enterotypes [19], with a high proportion of the North-Central population associated with enterotype-2 (73.5%) driven by *Prevotella* compared to the South India population sample (54% enterotype-2) where enterotype-1 (30.3%) driven by *Bacteroides* and enterotype-3 (16.1%) driven by *Ruminococcus* were also prevalent. This study also showed that species belonging to the genera *Bacteroides*, *Alistipes*, *Clostridium*, and *Ruminococcus* were relatively depleted in the Indian population compared to China, Denmark and the USA, whereas *Prevotella*, *Mitsuokella*, *Dialister*, *Megasphaera*, and *Lactobacillus* were the major drivers in separating Indian samples from other populations. In our study we found a predominance of enterotype-2 driven by *Prevotella*, with only a small number of individuals conforming to enterotype-1 driven by *Bacteroides* and a lack of enterotype-3 driven by *Ruminococcus*. High *Faecalibacterium* abundance was generally associated with high *Prevotella*. Bacterial taxa (e.g. *Dialister*, *Megasphaera*, *Mitsuokella*, *Lactobacillus*) previously shown to be characteristic of Indian gut microbiomes [20, 24, 46] occurred at <1% in our study population. Hence this population from North-East India does not fall within the criteria identified by these other studies to be characteristic of Indian populations. One reason for differences between our study and other in India could be differences in collection and processing of samples. Importantly here, there were no differences in the collection and processing of samples from VL cases and endemic control subjects that would influence our search for differences due to VL status.

One feature of the bacterial microbiota identified in our study was a high abundance of *Escherichia-Shigella* (9.1% aggregate abundance) in a subset of individuals that could define a different enterotype or sub-enterotype in our population compared to other Indian studies. This could reflect socio-economic status of residents of this region of India, together with higher prevalence of enteric diseases [47]. In a recent study of 475 patients with acute gastroenteritis Castaño-Rodríguez and colleagues [45] found that patient microbiomes clustered into three putative enterotypes, dominated by *Bacteroides*, *Escherichia-Shigella*, or *Faecalibacterium*. The novel *Escherichia-Shigella*-dominated enterotype found in a subset of patients was predicted to be more pro-inflammatory, and included the presence of *Enterobacteriaceae*, *Streptococcaceae*, and *Veillonellaceae*. In our study high *Escherichia-Shigella* abundance appeared to

be associated with an overall dysbiosis of the gut bacterial flora. Our study was also uniquely able to look at the interaction between prokaryotic and eukaryotic microflora in the gut. Interestingly, *Escherichia-Shigella* was the major bacterial taxon driving the effect of eukaryotic *Blastocystis* abundance on bacterial beta diversity measures, with a negative correlation between *Blastocystis* and *Escherichia-Shigella* log₁₀ counts. This might suggest that each out-competes the other and/or may reflect other influences of *Blastocystis* on gut microbial homeostasis (cf. below). Notably, however, there was no evidence for a difference in *Escherichia-Shigella* or other Enterobacteriaceae between VL cases and EC that might support a role for translocation across the gut of LPS from these bacteria in pathogenesis of VL as has previously been proposed [4, 5].

As alluded to above, our study was the first to employ 18S rRNA analysis of eukaryotic gut microflora in parallel with analysis of prokaryotic 16S rRNA sequences. Although overall measures of alpha and beta diversity for prokaryotic or eukaryotic microflora did not differ significantly between VL cases and EC, differences were observed at the level of individual bacterial or protozoan taxa. Specifically, using ANCOM we determined that EC had significantly higher levels of Ruminococcaceae UCG- 014 (Family: Ruminococcaceae) and Gastranaerophilales_ uncultured bacterium (Phylum: Cyanobacteria; Class: Melainabacteria; Order Gastranaerophilales) than VL cases. Both of these species are at relatively low aggregate relative abundance in our study population. However, previous studies in mice have shown that Ruminococcaceae UCG-014 is significantly reduced along with weight loss resulting from an intermittent fasting-mimicking diet [48]. This could reflect the effect of VL on cachexia and concomitant weight loss. Other research has noted perturbations of Ruminococcaceae UCG- 014 in a murine model of inflammatory bowel disease [49] and in mice fed on low versus high salt diets [50]. The only published data on uncultured *Gastranaerophilales* spp. found an increase with aging in mice, which the authors suggested pointed to diminished antimicrobial defence with age [51]. Larger sample sizes and further research will be required to determine the functional significance, if any, of the reduced abundance of Ruminococcaceae UCG- 014 and uncultured Gastranaerophilales in VL cases compared to EC.

A more prominent effect was the higher abundance of *P. hominis* in VL cases compared to EC. *P. hominis* is an anaerobic flagellated protist that colonizes the large intestine of a variety of mammals, including dogs, water buffalo, cattle, nonhuman primates, and humans [52, 53]. A recent epidemiological survey of faecal samples from northern China showed that 27% of dogs, 4% of human, and 47% of nonhuman primates carried *P. hominis* [52], with multiple authors suggesting that zoonotic transmission poses significant risk [52, 54, 55]. *P. hominis* has been associated with chronic diarrhoea and/or inflammatory bowel disease in animals [56], nonhuman primates [57] and humans [58], and is found in 14% of pregnant women in Papua New Guinea. In our study affected VL cases had higher abundance of *P. hominis* than EC. Due to their immunological status VL cases may be more vulnerable to zoonotic transmission in the context of close association between humans, dogs and cattle/buffalo in this poor rural region of North-East India.

Independently of VL *versus* EC status, we observed high prevalence of *Blastocystis* spp. in our study population. Studies in a rural region of Brazil have also observed high prevalence of *Blastocystis*, with some evidence for genetic heterogeneity of *Blastocystis* subtypes within the population [59]. The two main genus-level taxa observed in our study, *Blastocystis* sp. MJ99-568 and *Blastocystis* ambiguous taxa, had similar effects on bacterial alpha and beta diversity measures. This included the effect of *Escherichia-Shigella* in driving the influence of *Blastocystis* on low bacterial beta-diversity, and the negative correlation between *Blastocystis* and *Escherichia-Shigella* counts as noted above. Using ANCOM we also determined that high *Blastocystis* was associated with low *Bacteroides* and with high *Clostridiales vadinBB60*. Given the

former, the high prevalence of *Blastocystis* in our population could account for the low representation of bacterial enterotype-1 normally associated with *Bacteroides*. Within the gut *Bacteroides* are generally considered as friendly commensals [60]. Carbohydrate fermentation by *Bacteroides* produces volatile fatty acids that are absorbed through the large intestine and contribute to the host's daily energy requirement [61]. High *Blastocystis* appears to put *Bacteroides* at a disadvantage thus influencing gut flora homeostasis. Another aspect of the dysbiosis associated with high *Blastocystis* is its association with high *Clostridiales vadinBB60*. Studies in mice have shown that abundance of *Clostridiales vadinBB60* is influenced by manipulation of meat protein in the diet and plays a part along with other microbiota in bidirectional signalling between gut and brain to regulate body metabolism [62]. However, nothing is known concerning pathogenicity of this species. In summary, whilst high *Blastocystis* was negatively correlated with *Escherichia-Shigella*, it likely contributes to overall dysbiosis of gut microflora in this North-East Indian population particularly to the low prevalence of enterotype-1 normally characterised by high *Bacteroides*.

Finally, an important aim of our study was to determine whether helminths could be detected using 18S rRNA analysis. This analysis yielded aggregate abundances of 0.78% for Ascaridida correlating with *Ascaris* identified microscopically, 0.77% for Rhabditida correlating with *Strongyloides* (hookworm) by microscopy, and 0.19% Cyclophyllidea correlating with *Hymenolepis* tapeworm by microscopy. The only apparent failure was detection of *Trichuris* using 18S rRNA in participants for whom small numbers of eggs had been identified in the stool samples, although we were able to identify Trichocephalida amplicons in one study participant negative by microscopy. Our inability to identify all *Trichuris*-infected individuals could also reflect loss of power in the 18S rRNA sample due to the presence of a large degree of contaminating bacterial 16S rRNA amplification caused by our demonstration that the EMP primers were not eukaryote-specific as originally reported. This has now been corrected on the EMP web-site. The use of eukaryote-specific 18S rRNA primers in future studies should allow more efficient detection of eukaryotic taxa in human faecal DNA, including the full range of soil transmitted helminth species. Thus, while the sample size employed in this scoping study was underpowered to determine any association between soil transmitted gut helminths and VL, our approach combining analysis of 16S rRNA and 18S rRNA analysis should work well in future studies designed to determine the interplay between gut microbiota and susceptibility to infectious diseases such as VL. Such studies may benefit initially by combining amplicon sequencing with a qPCR panel for eukaryotes which would also provide information on parasite load.

In conclusion, this scoping study has provided novel data on prokaryotic and eukaryotic gut microbiota for individuals living in this region of North-East India endemic for VL. Interesting dysbioses have been observed when considering the influence of putative pathogenic prokaryotic and eukaryotic species on global measures of bacterial diversity in the gut, and some important associations seen between VL and the eukaryotic pathogen *P. hominis*. Our study provides useful baseline data upon which to develop a broader analysis of pathogenic enteric microflora and their influence on gut microbial health and NTDs generally.

Supporting information

S1 Checklist. STrengthening the Reporting of OBServational studies in Epidemiology (STROBE) checklist. Provides details of how our study complies with guidelines for epidemiological case-control studies set down by the STROBE consortium of international epidemiologists, methodologists, statisticians, researchers and journal editors. (PDF)

S1 Table. Metadata for endemic control (EC) and visceral leishmaniasis (VL) individuals contributing to the study.

(PDF)

S2 Table. Genus-level raw counts for 16S rRNA analysis.

(XLSX)

S3 Table. Genus-level raw counts for 18S rRNA analysis after filtering out a small number of human 18S rRNA sequences and the cross-contaminating bacterial 16S rRNA sequences.

(XLSX)

S1 Fig. Rarefaction plots based on Shannon's phylogenetic alpha diversity index for (a) 16S rRNA sequence data and (b) 18S rRNA sequence data (after filtering out contaminating bacterial sequences).

(PDF)

S2 Fig. Bar plots for relative abundance of taxa for 18S rRNA data prior to filtering out contaminating bacterial species identified as unclassified "Eukaryota". Taxa ordered by relative abundance of the contaminant "Eukaryota". (a) VL case and EC samples. (b) Negative (MD labels) and positive (MOCK labels) controls. Negative controls indicate a small degree of laboratory cross-contamination of taxa identified in experimental samples. (c) Bar plot for 16S taxa that were contaminants in the negative control samples, colour key to relative abundances as for main Fig 1(a). Note that these percentages are relative to the much smaller number of processed reads for negative controls (9573 ± 9294 compared to $71,123 \pm 22,122$ for positive mock controls and $91,923 \pm 69,706$ for VL cases and EC samples). The diversity within the samples was equivalent to VL cases and EC, indicating that they were likely due to cross-contamination in the endemic laboratory. The absolute numbers of reads were unlikely to have influenced the data and conclusions drawn for experimental samples.

(PDF)

S3 Fig. Influence of age and gender on 16S rRNA-determined prokaryotic microbial profiles. (a) and (b) bar plots for relative abundance of taxa by age and sex, respectively; colour key to relative abundances as for main Fig 1(a). (c) to (f) comparisons of species richness (number of ASVs) and alpha diversity measures (as labelled). (g) PCoA plots (left to right: PC1xPC2; PC1xPC3; PC2xPC3) for weighted UniFrac beta diversity by gender (see key). (h) Three-dimensional PCoA plot for weighted UniFrac beta diversity by age (see key for continuous scale).

(PDF)

S4 Fig. Influence of age and gender on 18S rRNA-determined eukaryotic microbial profiles. (a) and (b) bar plots for relative abundance of taxa by age and sex, respectively; colour key to relative abundances as for main Fig 1(a). (c) to (f) comparisons of species richness (number of ASVs) and alpha diversity measures (as labelled). (g) PCoA plots (left to right: PC1xPC2; PC1xPC3; PC2xPC3) for weighted UniFrac beta diversity by gender (see key). (h) Three-dimensional PCoA plot for weighted UniFrac beta diversity by age (see key for continuous scale).

(PDF)

S5 Fig. Influence of micro-geography on 16S rRNA-determined prokaryotic microbial profiles. (a) map of the district of Muzaffarpur showing subdistricts or block, annotated with the

numbers of samples collected from different blocks. Listed below the map are numbers of samples from blocks outside of Muzaffarpur. (b) bar plots for relative abundance of taxa by district; colour key to relative abundances of taxa as for main Fig 1(a). (c) to (f) comparisons of species richness (number of ASVs) and alpha diversity measures (as labelled) by blocks with $N \geq 4$. (g) PCoA plots (left to right: PC1xPC2; PC1xPC3; PC2xPC3) for weighted UniFrac beta diversity by colour-coded blocks.

(PDF)

S6 Fig. Influence of micro-geography on 18S rRNA-determined prokaryotic microbial profiles. (a) map of the district of Muzaffarpur showing subdistricts or block, annotated with the numbers of samples collected from different blocks. Listed below the map are numbers of samples from blocks outside of Muzaffarpur. (b) bar plots for relative abundance of taxa by district; colour key to relative abundances of taxa as for main Fig 2(a). (c) to (f) comparisons of species richness (number of ASVs) and alpha diversity measures (as labelled) by blocks with $N \geq 3$. (g) PCoA plots (left to right: PC1xPC2; PC1xPC3; PC2xPC3) for Bray's dissimilarity beta diversity by colour-coded blocks.

(PDF)

Acknowledgments

We would like to thank the hospital staffs at Kala-azar Medical Research Centre, Muzaffarpur for their assistance in the collection of samples and all research scholars of Infectious Disease Research Laboratory, Banaras Hindu University for their kind help during the study.

Author Contributions

Conceptualization: Rachael Lappan, Cajsa Classon, Om Prakash Singh, Jenefer M. Blackwell.

Data curation: Rachael Lappan, Cajsa Classon, Om Prakash Singh.

Formal analysis: Rachael Lappan, Ricardo V. de Almeida, Jenefer M. Blackwell.

Funding acquisition: Cajsa Classon, Shyam Sundar, Jenefer M. Blackwell.

Investigation: Rachael Lappan, Cajsa Classon, Shashi Kumar, Om Prakash Singh, Jaya Chakravarty, Poonam Kumari, Sangeeta Kansal.

Methodology: Rachael Lappan, Cajsa Classon, Jenefer M. Blackwell.

Project administration: Cajsa Classon, Om Prakash Singh, Jaya Chakravarty, Sangeeta Kansal.

Resources: Shyam Sundar, Jenefer M. Blackwell.

Software: Rachael Lappan.

Supervision: Shyam Sundar, Jenefer M. Blackwell.

Validation: Rachael Lappan, Cajsa Classon, Jenefer M. Blackwell.

Visualization: Rachael Lappan, Jenefer M. Blackwell.

Writing – original draft: Jenefer M. Blackwell.

Writing – review & editing: Rachael Lappan, Cajsa Classon, Jenefer M. Blackwell.

References

1. Alvar J, Velez ID, Bern C, Herrero M, Desjeux P, Cano J, et al. Leishmaniasis worldwide and global estimates of its incidence. *PLoS ONE*. 2012; 7(5):e35671. Epub 2012/06/14. <https://doi.org/10.1371/journal.pone.0035671> PMID: 22693548.
2. Murray HW, Berman JD, Davies CR, Saravia NG. Advances in leishmaniasis. *Lancet*. 2005; 366(9496):1561–77. [https://doi.org/10.1016/S0140-6736\(05\)67629-5](https://doi.org/10.1016/S0140-6736(05)67629-5) PMID: 16257344.
3. Clemente JC, Manasson J, Scher JU. The role of the gut microbiome in systemic inflammatory disease. *BMJ*. 2018; 360:j5145. Epub 2018/01/10. <https://doi.org/10.1136/bmj.j5145> PMID: 29311119.
4. Santos-Oliveira JR, Regis EG, Giacoia-Gripp CB, Valverde JG, Alexandrino-de-Oliveira P, Lindoso JA, et al. Microbial translocation induces an intense proinflammatory response in patients with visceral leishmaniasis and HIV type 1 coinfection. *J Infect Dis*. 2013; 208(1):57–66. Epub 2013/03/30. <https://doi.org/10.1093/infdis/jit135> PMID: 23539743.
5. Santos-Oliveira JR, Regis EG, Leal CR, Cunha RV, Bozza PT, Da-Cruz AM. Evidence that lipopolysaccharide may contribute to the cytokine storm and cellular activation in patients with visceral leishmaniasis. *PLoS Negl Trop Dis*. 2011; 5(7):e1198. Epub 2011/07/19. <https://doi.org/10.1371/journal.pntd.0001198> PMID: 21765960.
6. Zaiss MM, Rapin A, Lebon L, Dubey LK, Mosconi I, Sarter K, et al. The Intestinal Microbiota Contributes to the Ability of Helminths to Modulate Allergic Inflammation. *Immunity*. 2015; 43(5):998–1010. <https://doi.org/10.1016/j.immuni.2015.09.012> PMID: 26522986.
7. Hassan MF, Zhang Y, Engwerda CR, Kaye PM, Sharp H, Bickle QD. The *Schistosoma mansoni* hepatic egg granuloma provides a favorable microenvironment for sustained growth of *Leishmania donovani*. *Am J Pathol*. 2006; 169(3):943–53. <https://doi.org/10.2353/ajpath.2006.051319> PMID: 16936268.
8. Classon C, Feng X, Eidsmo L, Nylen S. Intestinal Nematode Infection Exacerbates Experimental Visceral Leishmaniasis. *Parasite Immunol*. 2019:e12618. Epub 2019/02/12. <https://doi.org/10.1111/pim.12618> PMID: 30742324.
9. Murthy PK, Dixit S, Gaur RL, Kumar R, Sahoo MK, Shakya N, et al. Influence of *Brugia malayi* life stages and BmAFII fraction on experimental *Leishmania donovani* infection in hamsters. *Acta Trop*. 2008; 106(2):81–9. <https://doi.org/10.1016/j.actatropica.2008.01.007> PMID: 18329620.
10. Azeredo-Coutinho RB, Pimentel MI, Zanini GM, Madeira MF, Cataldo JI, Schubach AO, et al. Intestinal helminth coinfection is associated with mucosal lesions and poor response to therapy in American tegumentary leishmaniasis. *Acta Trop*. 2016; 154:42–9. <https://doi.org/10.1016/j.actatropica.2015.10.015> PMID: 26519200.
11. O'Neal SE, Guimaraes LH, Machado PR, Alcantara L, Morgan DJ, Passos S, et al. Influence of helminth infections on the clinical course of and immune response to *Leishmania braziliensis* cutaneous leishmaniasis. *J Infect Dis*. 2007; 195(1):142–8. Epub 2006/12/08. <https://doi.org/10.1086/509808> PMID: 17152018.
12. Newlove T, Guimaraes LH, Morgan DJ, Alcantara L, Glesby MJ, Carvalho EM, et al. Anthelmintic therapy and antimony in cutaneous leishmaniasis: a randomized, double-blind, placebo-controlled trial in patients co-infected with helminths and *Leishmania braziliensis*. *Am J Trop Med Hyg*. 2011; 84(4):551–5. Epub 2011/04/05. <https://doi.org/10.4269/ajtmh.2011.10-0423> PMID: 21460008.
13. Llewellyn S, Inpankaew T, Nery SV, Gray DJ, Verweij JJ, Clements AC, et al. Application of a Multiplex Quantitative PCR to Assess Prevalence and Intensity Of Intestinal Parasite Infections in a Controlled Clinical Trial. *PLoS Negl Trop Dis*. 2016; 10(1):e0004380. Epub 2016/01/29. <https://doi.org/10.1371/journal.pntd.0004380> PMID: 26820626.
14. Maji A, Misra R, Dhakan DB, Gupta V, Mahato NK, Saxena R, et al. Gut microbiome contributes to impairment of immunity in pulmonary tuberculosis patients by alteration of butyrate and propionate producers. *Environ Microbiol*. 2018; 20(1):402–19. Epub 2018/01/13. <https://doi.org/10.1111/1462-2920.14015> PMID: 29322681.
15. Lee SC, Tang MS, Lim YA, Choy SH, Kurtz ZD, Cox LM, et al. Helminth colonization is associated with increased diversity of the gut microbiota. *PLoS Negl Trop Dis*. 2014; 8(5):e2880. <https://doi.org/10.1371/journal.pntd.0002880> PMID: 24851867.
16. Kreisinger J, Bastien G, Hauffe HC, Marchesi J, Perkins SE. Interactions between multiple helminths and the gut microbiota in wild rodents. *Philos Trans R Soc Lond B Biol Sci*. 2015; 370(1675). <https://doi.org/10.1098/rstb.2014.0295> PMID: 26150661.
17. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. *Nature*. 2012; 486(7402):222–7. Epub 2012/06/16. <https://doi.org/10.1038/nature11053> PMID: 22699611.

18. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. *Nature*. 2007; 449(7164):804–10. Epub 2007/10/19. <https://doi.org/10.1038/nature06244> PMID: 17943116.
19. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, et al. Enterotypes of the human gut microbiome. *Nature*. 2011; 473(7346):174–80. Epub 2011/04/22. <https://doi.org/10.1038/nature09944> PMID: 21508958.
20. Kulkarni AS, Kumbhare SV, Dhotre DP, Shouche YS. Mining the Core Gut Microbiome from a Sample Indian Population. *Indian J Microbiol*. 2019; 59(1):90–5. Epub 2019/02/08. <https://doi.org/10.1007/s12088-018-0742-0> PMID: 30728635.
21. Marathe N, Shetty S, Lanjekar V, Ranade D, Shouche Y. Changes in human gut flora with age: an Indian familial study. *BMC Microbiol*. 2012; 12:222. Epub 2012/09/28. <https://doi.org/10.1186/1471-2180-12-222> PMID: 23013146.
22. Patil DP, Dhotre DP, Chavan SG, Sultan A, Jain DS, Lanjekar VB, et al. Molecular analysis of gut microbiota in obesity among Indian individuals. *J Biosci*. 2012; 37(4):647–57. Epub 2012/08/28. PMID: 22922190.
23. Tuikhar N, Keisam S, Labala RK, Imrat, Ramakrishnan P, Arunkumar MC, et al. Comparative analysis of the gut microbiota in centenarians and young adults shows a common signature across genotypically non-related populations. *Mech Ageing Dev*. 2019. Epub 2019/02/10. <https://doi.org/10.1016/j.mad.2019.02.001> PMID: 30738080.
24. Dhakan DB, Maji A, Sharma AK, Saxena R, Pulikkan J, Grace T, et al. The unique composition of Indian gut microbiome, gene catalogue and associated faecal metabolome deciphered using multi-omics approaches. *Gigascience*. 2019. Epub 2019/01/31. <https://doi.org/10.1093/gigascience/giz004> PMID: 30698687.
25. Tanaka R, Hino A, Tsai IJ, Palomares-Rius JE, Yoshida A, Ogura Y, et al. Assessment of helminth biodiversity in wild rats using 18S rDNA based metagenomics. *PLoS ONE*. 2014; 9(10):e110769. <https://doi.org/10.1371/journal.pone.0110769> PMID: 25340824.
26. Little MD. Laboratory diagnosis of worms and miscellaneous specimens. *Clin Lab Med*. 1991; 11(4):1041–50. Epub 1991/12/01. PMID: 1802521.
27. Yu Y, Lee C, Kim J, Hwang S. Group-specific primer and probe sets to detect methanogenic communities using quantitative real-time polymerase chain reaction. *Biotechnol Bioeng*. 2005; 89(6):670–9. Epub 2005/02/08. <https://doi.org/10.1002/bit.20347> PMID: 15696537.
28. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet C, Al-Ghalith GA, et al. QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science. *PeerJ Preprints*. 2018; 6:e27295v2. <https://doi.org/10.7287/peerj.preprints.27295v2>
29. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010; 7(5):335–6. <https://doi.org/10.1038/nmeth.f.303> PMID: 20383131.
30. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal*. 2011; 17(1):10–2. <https://doi.org/10.14806/ej.17.1.200>.
31. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods*. 2016; 13(7):581–3. Epub 2016/05/24. <https://doi.org/10.1038/nmeth.3869> PMID: 27214047.
32. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *NucleicAcidsRes*. 2013; 41(Database issue):D590–6. <https://doi.org/10.1093/nar/gks1219> PMID: 23193283.
33. Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, et al. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome*. 2018; 6(1):90. Epub 2018/05/19. <https://doi.org/10.1186/s40168-018-0470-z> PMID: 29773078.
34. Blaxter M, Mann J, Chapman T, Thomas F, Whitton C, Floyd R, et al. Defining operational taxonomic units using DNA barcode data. *Philos Trans R Soc Lond B Biol Sci*. 2005; 360(1462):1935–43. Epub 2005/10/11. <https://doi.org/10.1098/rstb.2005.1725> PMID: 16214751.
35. Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE*. 2010; 5(3):e9490. Epub 2010/03/13. <https://doi.org/10.1371/journal.pone.0009490> PMID: 20224823.
36. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013; 30(4):772–80. Epub 2013/01/19. <https://doi.org/10.1093/molbev/mst010> PMID: 23329690.
37. Shannon CE. A mathematical theory of communication. *The Bell System Technical Journal*. 1948; 27:379–423 and 623–56.

38. Faith DP. Conservation evaluation and phylogenetic diversity. *Biol Conserv.* 1992; 61:1–10.
39. Mulder CPH, Baseley-White E, Dimitrakopoulos PG, Hector A, Scherer-Lorenzen M, Schmid B. Species evenness and productivity in experimental plant communities. *Oikos.* 2004; 107:50–63.
40. Bray JR, Curtis J.T. An ordination of upland forest communities of southern Wisconsin. *Ecological Monographs.* 1957; 27:325–49.
41. Lozupone CA, Hamady M, Kelley ST, Knight R. Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Appl Environ Microbiol.* 2007; 73(5):1576–85. Epub 2007/01/16. <https://doi.org/10.1128/AEM.01996-06> PMID: 17220268.
42. Vazquez-Baeza Y, Pirrung M, Gonzalez A, Knight R. EMPERor: a tool for visualizing high-throughput microbial community data. *Gigascience.* 2013; 2(1):16. Epub 2013/11/28. <https://doi.org/10.1186/2047-217X-2-16> PMID: 24280061.
43. Mandal S, Van Treuren W, White RA, Eggesbo M, Knight R, Peddada SD. Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microb Ecol Health Dis.* 2015; 26:27663. Epub 2015/06/02. PMID: 26028277.
44. Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A, et al. Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome.* 2017; 5(1):27. Epub 2017/03/04. <https://doi.org/10.1186/s40168-017-0237-y> PMID: 28253908.
45. Castano-Rodriguez N, Underwood AP, Merif J, Riordan SM, Rawlinson WD, Mitchell HM, et al. Gut Microbiome Analysis Identifies Potential Etiological Factors in Acute Gastroenteritis. *Infect Immun.* 2018; 86(7). Epub 2018/04/25. <https://doi.org/10.1128/IAI.00060-18> PMID: 29685983.
46. Bhute S, Pande P, Shetty SA, Shelar R, Mane S, Kumbhare SV, et al. Molecular Characterization and Meta-Analysis of Gut Microbial Communities Illustrate Enrichment of Prevotella and Megasphaera in Indian Subjects. *Front Microbiol.* 2016; 7:660. Epub 2016/06/01. <https://doi.org/10.3389/fmicb.2016.00660> PMID: 27242691.
47. State Surveillance Unit SHS, Bihar. Annual Communicable Disease Surveillance Report, 2015 Bihar. http://16410013011:8091/idsp/annually/Annual_communicable_disease_surveillance_report-2015pdf 2015.
48. Wei S, Han R, Zhao J, Wang S, Huang M, Wang Y, et al. Intermittent administration of a fasting-mimicking diet intervenes in diabetes progression, restores beta cells and reconstructs gut microbiota in mice. *Nutr Metab (Lond).* 2018; 15:80. Epub 2018/11/28. <https://doi.org/10.1186/s12986-018-0318-3> PMID: 30479647.
49. Liu G, Yan W, Ding S, Jiang H, Ma Y, Wang H, et al. Effects of IRW and IQW on Oxidative Stress and Gut Microbiota in Dextran Sodium Sulfate-Induced Colitis. *Cellular physiology and biochemistry: international journal of experimental cellular physiology, biochemistry, and pharmacology.* 2018; 51(1):441–51. Epub 2018/11/20. <https://doi.org/10.1159/000495240> PMID: 30453297.
50. Wang C, Huang Z, Yu K, Ding R, Ye K, Dai C, et al. High-Salt Diet Has a Certain Impact on Protein Digestion and Gut Microbiota: A Sequencing and Proteome Combined Study. *Front Microbiol.* 2017; 8:1838. Epub 2017/10/17. <https://doi.org/10.3389/fmicb.2017.01838> PMID: 29033907.
51. van der Lugt B, Rusli F, Lute C, Lamprakis A, Salazar E, Boekschoten MV, et al. Integrative analysis of gut microbiota composition, host colonic gene expression and intraluminal metabolites in aging C57BL/6J mice. *Aging (Albany NY).* 2018; 10(5):930–50. Epub 2018/05/18. <https://doi.org/10.18632/aging.101439> PMID: 29769431.
52. Li WC, Ying M, Gong PT, Li JH, Yang J, Li H, et al. Pentatrichomonas hominis: prevalence and molecular characterization in humans, dogs, and monkeys in Northern China. *Parasitology research.* 2016; 115(2):569–74. Epub 2015/10/21. <https://doi.org/10.1007/s00436-015-4773-8> PMID: 26481488.
53. Kamaruddin M, Tokoro M, Rahman MM, Arayama S, Hidayati AP, Syafruddin D, et al. Molecular characterization of various trichomonad species isolated from humans and related mammals in Indonesia. *Korean J Parasitol.* 2014; 52(5):471–8. Epub 2014/10/30. <https://doi.org/10.3347/kjp.2014.52.5.471> PMID: 25352694.
54. Fang YK, Chien KY, Huang KY, Cheng WH, Ku FM, Lin R, et al. Responding to a Zoonotic Emergency with Multi-omics Research: Pentatrichomonas hominis Hydrogenosomal Protein Characterization with Use of RNA Sequencing and Proteomics. *Omics: a journal of integrative biology.* 2016; 20(11):662–9. Epub 2016/11/10. <https://doi.org/10.1089/omi.2016.0111> PMID: 27828770.
55. Maritz JM, Land KM, Carlton JM, Hirt RP. What is the importance of zoonotic trichomonads for human health? *Trends Parasitol.* 2014; 30(7):333–41. Epub 2014/06/22. <https://doi.org/10.1016/j.pt.2014.05.005> PMID: 24951156.
56. Bastos BF, Brener B, de Figueiredo MA, Leles D, Mendes-de-Almeida F. Pentatrichomonas hominis infection in two domestic cats with chronic diarrhea. *JFMS Open Rep.* 2018; 4(1):2055116918774959. Epub 2018/06/07. <https://doi.org/10.1177/2055116918774959> PMID: 29872537 with respect to the research, authorship, and/or publication of this article.

57. Laing ST, Merriam D, Shock BC, Mills S, Spinner A, Reader R, et al. Idiopathic Colitis in Rhesus Macaques Is Associated With Dysbiosis, Abundant Enterochromaffin Cells and Altered T-Cell Cytokine Expression. *Veterinary pathology*. 2018; 55(5):741–52. Epub 2018/06/23. <https://doi.org/10.1177/0300985818780449> PMID: 29929446.
58. Meloni D, Mantini C, Goustille J, Desoubreux G, Maakaroun-Vermesse Z, Chandenier J, et al. Molecular identification of *Pentatrichomonas hominis* in two patients with gastrointestinal symptoms. *Journal of clinical pathology*. 2011; 64(10):933–5. Epub 2011/05/10. <https://doi.org/10.1136/jcp.2011.089326> PMID: 21551465.
59. Barbosa CV, Barreto MM, Andrade RJ, Sodre F, d'Avila-Levy CM, Peralta JM, et al. Intestinal parasite infections in a rural community of Rio de Janeiro (Brazil): Prevalence and genetic diversity of Blastocystis subtypes. *PLoS ONE*. 2018; 13(3):e0193860. Epub 2018/03/10. <https://doi.org/10.1371/journal.pone.0193860> PMID: 29522552.
60. Wexler HM. Bacteroides: the good, the bad, and the nitty-gritty. *Clin Microbiol Rev*. 2007; 20(4):593–621. Epub 2007/10/16. <https://doi.org/10.1128/CMR.00008-07> PMID: 17934076.
61. Hooper LV, Midtvedt T, Gordon JI. How host-microbial interactions shape the nutrient environment of the mammalian intestine. *Annu Rev Nutr*. 2002; 22:283–307. Epub 2002/06/11. <https://doi.org/10.1146/annurev.nutr.22.011602.092259> PMID: 12055347.
62. Xie Y, Zhou G, Wang C, Xu X, Li C. Specific Microbiota Dynamically Regulate the Bidirectional Gut-Brain Axis Communications in Mice Fed Meat Protein Diets. *J Agric Food Chem*. 2019; 67(3):1003–17. Epub 2018/12/14. <https://doi.org/10.1021/acs.jafc.8b05654> PMID: 30541283.