THE MICROBIOME OF ACUTE BACTERIAL GASTROENTERITIS AND THE FUNCTIONAL ROLE OF INTESTINAL BACTERIOPHAGES

By

Brian Nohomovich

A DISSERTATION

Submitted to
Michigan State University
in partial fulfilment of the requirements
for the degree of

Microbiology and Molecular Genetics - Doctor of Philosophy

2019

ABSTRACT

THE MICROBIOME OF ACUTE BACTERIAL GASTROENTERITIS AND THE FUNCTIONAL ROLE OF INTESTINAL BACTERIOPHAGES

By

Brian Nohomovich

Acute gastroenteritis has a major disease burden worldwide. There are 2.3 billion cases of acute gastroenteritis worldwide each year that accounts for 8% of all deaths in children under the age of 5. In the United States, there are an estimated 179 to 375 million cases annually. Gastroenteritis can have acute and chronic effects on human health. Pathogens often are not identified in cases of acute gastroenteritis due in part to the wide range of causative agents and the difficulties with standard culturing practices. The advent of next-generation sequencing has allowed the study of the intestinal microbiome to detect alterations in the composition as specific disease signatures. There have been few studies on the microbiome of gastroenteritis, but none have to date have studied both the virome and bacteriome together. Through this combined analysis, a deeper understanding of gastroenteritis can be generated.

In this dissertation, the Microbiome (Virome and Bacteriome) of 79 cases and 125 member controls were examined. It was found that cases had lower diversity and richness in and increased abundances in Enterobacteriaceae. Additionally, associations with severe illness were made to a specific cluster of samples. Differential abundance analysis identified the involvement of both viruses and bacteria. Analysis of the same 79 cases in a recovery state (n=63), identified the changes that occur during and after infection. These changes agree with the case and control analysis. The functional aspects were analyzed of the viral communities. Three novel bacteriophages were isolated from stool samples and characterized. Two of the bacteriophages

were determined to be lysogenic and were found in 23 additional E. coli O157:H7 strains based on BLAST alignments. One of the lysogenic bacteriophages (PHG003), harbors an SbcC gene which is a predicted exonuclease but it's important to the host bacterium remains unknown. Additionally, a lytic bacteriophage (PHG001) was also isolated and exhibited a relatively broad host range and was incredibly virulent to E. coli O157:H7. Additionally, PHG001 exhibits a phage-antibiotic synergism with the use of ampicillin and mitomycin c. Either antibiotic with the bacteriophage exhibited a drastic reduction in bacteria growth. PHG001 also reduced shiga toxin expression compared to control levels.

For Maui and my boo

ACKNOWLEDGEMENTS

The work in this dissertation was started in the summer of 2013 on my first rotation project. This dissertation represents more than six years of active research and medical school requirements. The support that I have received has been immense and I could not have done it alone. I would like to thank Dr. Shannon Manning for supporting me in this endeavour. My committee (Dr. Andrea Amalfitano, Dr. Terrance Marsh, Dr. Arjun Krishnan, Dr. Sun) has facilitated my growth into a scientist. The MMG department has made this process fun throughI discussions at WiPS, GSW, or Tea at Three. The CMSE department and ICER have both been immensely helpful in helping me build the skillset and providing the resources necessary to complete the work. I think Spartan Innovations for giving me the opportunity to use my computational skillset in a different way. I would like to give a special appreciation to Kristen Parent and Jason Schrad for taking such wonderful bacteriophage micrographs. Heather Blankenship, thank you for the fun bioinformatic talks.

My family and friends have been incredibly supportive through this process and I thank them for providing an environment that was not science all the time. I give my deepest appreciation to Megan Shiroda. I am so grateful that we accomplished this feat together, it has been challenging, but we did it. I am looking forward to tackling the next big challenge with you. Thank you for always being there.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

KEY TO ABBREVIATIONS

AMP	Ampicillin
ANCOM	ANalysis of Composition Of Microbiomes
ANOVA	ANalysis Of VAriance
ARG	Antibiotic Resistance Genes
AUC	Area Under Curve
ATCC	American Type Culture Collection
BAM	Bacteriophage Adhering to Mucus
BLAST	Basic Local Alignment Search Tool
CDC	Centers for Disease Control and Prevention
CI	Confidence Interval
CFU	Colony Forming Units
EHEC	Enterohemorrhagic *E. coli*
ERIN	Enterics Research Investigational Network
FDA	Food and Drug Administration
FoodNet	Foodborne Disease Active Surveillance Network
HUS	Hemolytic Uremic Syndrome
LB	Luria Broth
MDHHS	Michigan Department of Health and Human Services
MDSS	Michigan Disease Surveillance System
MOI	Multiplicity of Infection
NCBI	National Center for Biotechnology Information

NGS	Next Generation Sequencing
NPV	Negative Predictive Value
OR	Odds Ratio
PCR	Polymerase Chain Reaction
PEG	Polyethylene Glycol
PCA	Principle Components Analysis
PFU	Plaque Forming Units
PHAST(ER)	PHAge Search Tool
PPV	Positive Predictive Value
PT	Phage Type
qRT-PCR	Real-Time Quantitative Reverse Transcription PCR
ROC	Reciever-Operator Curve
SEM	Scanning Electron Microscope
Stx	Shiga toxin
VLP	Virus-Like Particles
WGS	Whole Genome Sequencing
WHO	World Health Organization

CHAPTER 1

LITERATURE REVIEW: THE MICROBIOME OF GASTROENTERITIS

THE DISEASE BURDEN OF DIARRHEAL ILLNESS

Acute gastroenteritis (infectious diarrhea) is a significant health burden and is one of the most common illnesses requiring hospitalization globally (1). There are 2.3 billion cases of acute gastroenteritis (2). 1.3 million deaths annually occur as a result diarrhea and in 2016 it accounted for 8% of all deaths among children under the age of 5 years (3). Diarrheal illness contributes to one in eight deaths in children younger than five years.

With most infections occurring in developing countries (4), the estimated incidence of acute gastroenteritis in the United States ranges from 179 million (5) to 375 million cases (6), though this likely is an underestimation of the true incidence, as 50% of cases present without symptoms (7). There are 1.5 million office visits, 200,000 hospitalizations, and 300 deaths due to acute gastroenteritis of children in the United States (8). Studies have identified the causative agent of gastroenteritis in 2.4% to 32% of cases (9–11). A recent study found that 10% of hospitalized cases (n=196) tested positive for known gastroenteritis pathogens (12). In contrast, lower identification rates (1.5%) of the causative agent was identified when culturing for a subset of known pathogens in cases (13). Screening for a large number of pathogens increases the likelihood of identifying a causative agent in a case. However, extensive culturing for pathogen identification in cases of gastroenteritis is both cost and time prohibitive to perform. Furthermore, empiric management for most patients, which consists of nutritional support and the avoidance of antibiotics (except in select cases) (14, 15), will remain relatively unchanged despite a positive stool-culture result (9). Given these limitations, culturing is still the primary diagnostic test for laboratory diagnosis in gastroenteritis (16). Although whole-genome sequencing of isolated pathogens can be used to confirm identity and predict virulence and phenotypes based on genomic alignments, culturing is still necessary to evaluate important phenotypes such as antibiotic

susceptibility, serotyping, and the expression of specific virulence factors. Studies have attempted to improve the predictability of stool culturing by utilizing serum c-reactive protein (CRP) and stool white blood cell counts (13), with some suggesting the use of a scoring system along with clinical presentation, stool culture, and CRP to guide patient management (17).

The poor diagnostic yield and clinical utility of stool culturing is due in part to the great diversity of organisms that can potentially cause acute gastroenteritis and the impracticalities of directly culturing for each potential pathogen. The most common potential pathogens that can cause diarrhea include viruses such as *Norovirus*, *Astrovirus,* and *Rotavirus* (5). Bacterial pathogens include *Campylobacter jejuni*, *Escherichia coli* (*E. coli*), *Salmonella* spp., and *Shigella* spp., while protozoan pathogens include *Cryptosporidium* and *Giardia*. Helminths such as *Ascaris* and *Enterobius* are also common causative organisms for diarrhea (18). *Rotavirus*, *Cryptosporidium*, *Shigella*, and *E. coli* account for most of the disease burden globally in children (19). *E. coli* infections can further impact childhood development (20) and induce acute kidney injury (21).

Gastroenteritis has both acute and chronic indirect impacts on human health. An acute indirect effect involves immediate pathogen infection and resolution that as a result also leads to the expansion of *Enterobacteriaceae* in the gut (22). It can also lead to a chronic, inflammatory state in the gut that predisposes patients to post-infectious irritable bowel syndrome (IBS), or inflammatory bowel disease (IBD), with symptoms lasting up to 10 years after the infection (23, 24). In the year following a case of infectious gastroenteritis, individuals are 2.4 times more likely to develop IBD (25). An underlying mechanism predisposing to these chronic conditions has been proposed, and involves triggering of a divergent inflammatory response due to the initial infection (22). This response, which was observed for infections caused by adherent invasive *E. coli* (AIEC),

creates a selective environment for bacterial proliferation and prolonged inflammation (26). Defining significant alterations that occur in the human gut microbiome during bacterial infections, otherwise known as intestinal dysbiosis, can not only identify novel mechanisms contributing to several human diseases, but may also lead to the identification of novel therapeutic interventions.

THE MICROBIOME

Microbiome as a term was first used in 1952 (27) and referred to the entire ecological community of microbes and their interactions with the immediate environment. A more modern definition of microbiome refers to the collective genomes of the microbes in a respective environment (28). Both definitions are interchangeably used in contemporary research studies. The former focuses on the main microbial members in a community, while the latter integrates molecular genomics to infer the presence of several additional elements, such as metabolic functions, in these communities. There are many determinants of the composition of the microbiome. For instance, studies have shown that diet (29), antibiotics (30), genetics (31), age, and geography (32, 33) can all shape a respective microbiome. Diseases can influence microbiomes, which has been demonstrated for diabetes (34), obesity (35), cancer (36), IBD (37), HIV (38), rheumatoid arthritis (39), and gastroenteritis (40). Sequencing technology is commonly used to study the microbiome due to difficulties in culturing all microbes residing in a given community.

Characterizing the Microbiome

The first application of genomics for characterizing a microbial community occurred in 1986 with the use of vectors like Bacteria Artificial Chromosomes (41). In short, fragments of DNA present in a respective microbiome were subcloned and subjected to DNA sequencing techniques. The identification of unique DNA sequences, and the subsequent alignment of these

against other known genomes, allows analysis of the genomic architecture of the overall microbiome community. The advent of high throughput and automated sequencing technologies expanded the capacities of these studies, fostering more sophisticated methods to identify organisms via gene analysis. One example of this is called metataxonomics (42) a method that involves selective high throughput sequencing of single marker genes to identify and classify the organisms present in a given sample (43). Sequencing the ribosomal RNA (rRNA) genes, for instance, is useful for profiling the taxonomical composition of distinct microbial communities. Common marker genes include 16S rRNA for bacteria (43), 18S rRNA for eukaryotes (44), and the internal transcribed spacer (ITS) region of the ribosome for fungi identifications (45). Marker genes are well conserved and allow for differentiation to the species level (46) due to sequence variation in the hypervariable region of the target gene sequence. Quantitative Insights Into Microbial Ecology (47) or mothur (48) are examples of algorithms used for marker gene analysis and involve quality filtering, denoising (error correction), chimeric sequence removal, clustering of reads into operational taxonomic units (OTUs) and classification of OTUs utilizing a database such as the Ribosomal Database Project (RDP) (49), SILVA (50), or the now-defunct Greengenes (51). Use of these strategies have been instrumental in defining the bacteriome in environments, animals, and healthy and ill humans. Nonetheless, the use of marker genes has several significant limitations. Marker genes are imperfect; more than 50% of organisms are undetected with 16S rRNA amplicon sequencing (52). Additionally, viruses cannot be classified using this technology as they lack analogous universal conserved genes to serve as a unique identifier for an organism.

Metagenomics utilizes high-throughput, non-targeted DNA sequencing of the microbial genomes in an environment without targeting a particular marker gene (53). Sequencing produces short fragments of base pairs, a “read”, representing a portion of a genome. Metagenomic analysis

of a microbiome begins with quality-control. Numerous tools are available for quality-control analysis, and include Cutadapt (54), Trimmomatic (55), FastX-Toolkit (56), and BBtools (57). FastQC (58) can be used, while MultiQC (59) can merge individual reports into a single final report. Read alignment software such as Bowtie2 (60), BWA (61), or FastQ Screen (62) can be used to match reads against reference genomes (e.g., human) and to remove host sequences that were also sequenced during amplification. For larger datasets, digital normalization (63) with the Khmer package (64) can be used to reduce read redundancy and normalize coverage in samples, thereby making downstream analyses computationally cheaper.

After quality-control assessments, sequenced reads can be assembled into contiguous sequences (contigs) or classified directly. Direct classification of reads is sufficient to profile the microbial community from environments with related microbial populations that are well studied (i.e. the human gut). This approach allows for an assessment of all types of genomes including viruses and does not have the bacterial bias inherent with 16S rRNA gene analyses. Additionally, species diversity, richness and uniformity of each community can be evaluated for each profile (65). Reads can be directly mapped to curated pathogen databases of genes to provide insight about gene functions using the tools listed above. Novel pathogens (66, 67) have been discovered using these approaches, and in microbial environments that are relatively unexplored, assembly and binning of sequences can provide a qualitative assessment of the respective microbiome.

Deriving sequencing data and distilling the information into identification of specific organisms via single genome assembly is challenging, as assembly involves matching reads into longer contigs that can be used for downstream analyses (68). Two of the most common single genome assemblers are Velvet (69) and SPAdes (70). Although traditional assemblers assume uniform coverage across the genome to help resolve errors, metagenomics assembler such as Meta-

Velvet (71), MegaHit (72), IDBA-UD (73), and Meta-SPAdes (74) relax this assumption. Additionally, metagenomics involves a mixed population of microbes at varying abundances and uneven sequencing depth, which needs to be accounted for by metagenomic assemblers. Metatranscriptomics is an alternative strategy to study functional metagenomics and attempts to capture all the RNA in a sample, which represents all genes that were transcribed (86). Regardless of the approach taken, annotation of the sequences must be performed before this analysis can be started.

The functional capacity of an organism can be investigated directly once the nearly complete genome is available (78). Genes related to virulence or function can directly be extracted from MAGs and analyzed for multiples purposes including constructing phylogenetic trees to elucidate evolutionary relationships and diversity within a sample. Also, functional metagenomics can infer the translated product of identified gene sequences either through inference using software like Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) (84) or direct expression of the gene in a vector (85).

Metagenomic classification and sequence profiling can be performed on reads, assemblies, or MAGs. The Basic Local Alignment Search Tool (BLAST) (87), which is available in Genbank® (88) via the National Center for Biotechnology Information (NCB), is the traditional methodology used for aligning sequences. However, due to increases in both the database size and number of sequenced datasets, it has become computationally impractical to utilize BLAST alignments. Despite this, BLAST remains the most sensitive software (89) even though additional tools have been developed to expedite the classification process. Sequence classification techniques fall under one of the following categories: alignment-based approach Bowtie2 (60); kmer-based alignment with Kraken (90) or Clark (91); aligning translated nucleotides to a protein database by sequence

with Diamond (92) or by kmer with Kaiju (93); alignment of marker genes with PhyloPhlAn (94); or alignment of minhash signatures (95) with MASH (96) or Sourmash (97). The selection of the toolset to be used should reflect both the research goals and computational facilities available.

Assemblies can be used to derive complete or near-complete genomes from metagenomes called metagenome-assembled genomes (MAGs). The process of creating a MAG involves binning of contigs to identify individual genomes within metagenomic samples (79). To do this, binned contigs are quality-controlled with GroopM CheckM (80), MaxBin 2.0 (81), and Metabat 2 (82). GroopM (83) infers the population genomes by coverage of assembled contigs, while CheckM measures the completeness and contamination of a MAG with the use of single-marker genes. Moreover, MaxBin 2.0 utilizes an expectation-maximation algorithm to optimize the number of bins, and Metabat 2 merges contigs into MAGs. This tool is particularly useful for the identification of new species as was demonstrated in a prior study of microbial communities within hot springs, which identified 36 MAGs, some representing taxonomically underrepresented groups like archaea (78) or in a cattle rumen which uncovered 913 novel species of bacteria (79). Despite these difficulties, metagenomic assemblies have been used to identify a novel bacteriophage (75). This bacteriophage is a member of the most abundant bacteriophage family in the human gastrointestinal tract (76) and was recently isolated and cultured (77).

The Bacteriome

The microbiome is an umbrella term to represent all microorganisms residing in a given environment. The microbiome consists of other "-omes" (28) that are specific for bacteria, fungi, archaea, and viruses. Studies on the bacterial component of the microbiome, "the bacteriome," have successfully documented the bacterial communities present via *16S rRNA* sequencing, metagenomics (98), and culturing (99). The bacteria of the microbiome have been studied in

animals (100), at various sites in the human body (101), and in many human disease states (102–105). Efforts have been employed to translate knowledge of a respective microbiome to clinical laboratory tests that can be utilized to improve patient care while classifying the common pathogenic and non-pathogenic taxa (106).

Much work has been done relative to identifying crucial bacterial members of the distal gut microbiome in both health and disease states. Firmicutes and Bacteroides have been found to represent the most dominant phyla in the human colon (35), and alterations in phyla abundance have been associated with obesity (34, 35, 107, 108). *Bacteroidetes* and *Prevotella*, both genera belonging to Bacteroides, are commonly found within the human gut. Enterotypes have been proposed that classify an individual's intestinal microbiome based on the dominant genera of either *Bacteroidetes* or *Prevotella*, which were suggested to have an antagonistic relationship (109). Studies on diet (29) have shown that western-type diets have a high-abundance of *Bacteroidetes*, with 40-60% of an individual's microbiome being comprised of this genus (110). A meta-analysis of diet studies further identified that *Prevotella* and *Bacteriodetes* represent the most significant percentage of a healthy person's fecal microbiome and could be utilized as a marker for diet, with both genera representing biomarkers of diet and lifestyle (111). A study in germ-free mice also suggests competition between *Bacteriodetes thetaiotaomicron* and *Prevotella copri*, which occurred due to increased fiber intake (112). The *Prevotella* population, in particular, may be important for plant glycan digestion. Indeed, prior studies have linked *Prevotella* abundance to plant-based or Mediterranean diets even though its specific metabolic niche remains ill-defined (113); this could be due in part to its high degree of diversity (114). In a mouse model using twin microbiota discordant for obesity (107), it was found that the phenotype correlated with microbiota profiles. The obese phenotype in mice was associated with branch-chain amino acid metabolism

whereas the lean phenotype was associated with small-chain fatty acid fermentation. Small-chain fatty acids, such as butyrate, have been shown to decrease insulin sensitivity and energy expenditure in mice (115) and regulate gut hormones to promote a lean phenotype (116). *Bacteroidetes* abundance was associated with the lean phenotype in mice (107) and is probably due to being and is a butyrate producer (117). Methanogens have also been associated with a leaner body habitus (118). *Chrisensenella* is the most dominant methanogen family in the human gut (119) and is the most heritable bacteria family in repeat twin studies (120, 121). Methanogens can reduce hydrogen to methane, which promotes the growth of anaerobic bacteria (122).

Additional populations of bacteria important for human gut health include *Veillonella* (123), and *Bacteroidetes* (124) that can both metabolize bile acids, which is vital for dietary intake of fats and fat-soluble vitamins. *Akkermansia* is involved in mucin degradation (125) which can lead to mucosal degradation (126). *Bifidobacterium* can promote health by the breakdown of sugars (127), which can then be cross-fed to other microbiota, like small-chain fatty acid producers (128). *Odoribacter* (129) and *Roseburia* (130) can produce small-chain fatty acids which have been shown to have anti-inflammatory effects (131). *Faecalbacterium* produces anti-inflammatory proteins that reduce inflammation in the gastrointestinal tract. The presence of these anti-inflammatory proteins has been found deficient in Crohn's disease patients and, therefore, could play a role in reducing colitis in mouse models (132). Similarly, *Enterococcus* represents a group of common commensals that can produce bacteriocins with antimicrobial properties (133); some members of *Enterococcus*, however, are opportunistic pathogens and can cause infections. Members of the phylum, Proteobacteria, are increased in abundance in disease (134). For example, *Escherichia*, a genus of Proteobacteria, has been associated with IBD (135, 136), gastroenteritis (40), and colorectal cancer (135, 137). By contrast, other genera within Proteobacteria can exhibit

anti-inflammatory properties. *Acinetobacter*,a genus of Proteobacteria, has been shown to directly induce T cell differentiation *in vitro* and downregulate helper T cells (138). Like *Enterococcus*, some members such as *Acinetobacter baumannii*, are opportunistic pathogens capable of causing human infections. *Alistipes* typically represent commensals found in lower abundance and have been associated with plant-based diets (139, 140); but, it was associated with abdominal pain in pediatric patients (141).

The Virome

Viruses can also affect both bacteria and humans. Viruses found within microbial communities examined by metagenomics represent the "virome". As some of the viruses within the virome are bacteriophages, ie: viruses that infect bacteria, these microbial populations are commonly referred to as the "phageome". Both play essential roles in shaping bacterial communities in any environment. Prior studies have classified viral communities using the multiple-displacement amplification (142), direct isolation of viruses with sequencing (143, 144), and viral genome identification in metagenomes (145). Through these studies it has become apparent that viral databases are sparse (146) with many of the genomes of isolated viruses not aligning to known viruses (143, 147–150). Importantly, assemblies of reads from metagenomes of isolated viruses have resulted in less than 2% of the sequences getting annotated taxonomically (151). This is in stark contrast to bacterial databases that can achieve greater than 90% annotation of the diversity in sequencing reads down to the species level in the human gastrointestinal tract (152). Viruses mutate more frequently than their hosts, and even if there exists an entry in the database, they still might not get classified. Despite the incompleteness of viral databases and the difficulty with the approaches in studying the virome, studies of viral communities have provided great insight into ecology and human health.

The Phageome

Most studies have focused on healthy individuals to describe the bacteriophage component of the virome, "the phageome". Studies have shown that the intestinal phageome rapidly changes in the first weeks of life (153) and through childhood into adulthood (154). The variation of the virome between individuals is high, but diversity within a person is low (143). Additionally, diet (147), antibiotics (155), and chronic diseases such as HIV (38), IBD (156), and colorectal cancer (157) have been shown to impact the phageome directly. Such chronic insults have been shown to contribute to rapid phageome evolution (150) driven mostly by temperate bacteriophages (144, 158). Temperate bacteriophages, also called lysogenic phage, are bacteriophages that are incorporated in the bacterial genome as a prophage. The prophage is maintained and replicated alongside the host bacterial genome but can enter a lytic state, releasing progeny into the environment, which increases the abundance of bacteriophage present.

Studies that have examined interactions between bacteriophage and their host bacteria in aquatic environments (159) have proposed a "kill-the-winner" (KTW) model for microbial ecosystems (160), an expansion Lotka-Volterra cycling model for predator-prey relationships (161). In the KTW model, latent prophages replicate in proportion to the abundance of its host bacteria, which results in a stable bacterial population (162). For instance, an outgrowth of a bacterial population could be lysed by increased replication of prophages present in its genome. Similar dynamics have been observed in the human gut as well. A prophage of *Enterococcus faecalis* expressed in the presence of amino acids, for example, resulted in a reduction of the bacterial host (163). Indeed, nutrient availability has long been established as a predictor of prophage induction (164–168).

Based on genomic alignments performed in one study, it was demonstrated that bacteriophages were capable of infecting several hosts within the oral microbiome by identification of the same prophages in different bacterial species. (169). There is evidence for cross-infectivity, or the ability of one bacteriophage to infect multiple hosts, outside of metagenomics. For example, a Myoviridae bacteriophage could infect Shiga toxin-producing *E. coli* (STEC) O157:H7 and *Salmonella* spp., two common but genetically distinct members of Proteobacteria (170). Tunavirinae, however, is a subfamily of bacteriophage that has high specificity for STEC O157:H7 but little infectivity for non-O157 STEC (171), demonstrating that variation in infectivity can also occur among members of the same species. KTW does not fully explain this finding of cross-infectivity in bacteriophages as it typically models a single bacteriophage-bacterium relationship.

THE IMPACT OF BACTERIOPHAGE ON MICROBIOTA

A bacteriophage that has multiple hosts would have an increased chance of replicating and persisting in a given environment. The initial process of a bacteriophage infecting a bacterium is due to the presence of a receptor on the host bacterium. Detailed analysis is needed to identify critical receptors that bacteriophage targets to better define their relationship with specific members of the microbiota (172). Bacteriophage can directly impact host bacterial populations by altering transcription (173) and providing them with beneficial genes such as toxins (174, 175) or antibiotic resistance genes (143, 147, 176, 177) that facilitate survival in different environments. Few studies have directly studied both the virome and bacteriome simultaneously but have provided great insight into ecology. Studies of monozygotic twins have demonstrated that bacteriophage populations within the microbiome can directly shape bacterial diversity and that bacteriome abundance is inversely correlated with virome abundance (178). The inverse

correlation between the abundance of viruses and bacteria has been observed in other studies (38, 156, 157). Furthermore, mouse models have shown that bacteriophage can directly impact the resident microbes (179), while mucosal models of confluent cell layers demonstrated that bacteriophage could transcytose across the mucosal surface (180). Similarly, another study found that bacteriophage in a mucin matrix could prevent pathogen colonization (181) while accumulating in the mucosa to a concentration 10x higher than the bacterial concentration. (181) This finding is in stark contrast to the bacteria to bacteriophage ratios that have been described in feces, which are generally 10:1 to 1:1 (143, 147, 158). Mucosal surfaces are common infection sites of invading pathogens, and bacteriophages are frequently found at these sites (181). Additionally, small intestine bacteriophage adhere to the KTW model by preserving bacterial diversity (182), in contrast to bacteriophage of the large intestine that fail to preserve bacterial diversity (143).

The "Piggyback-the-winner" (Figure 1.1) theory, attempts to reconcile some of the issues associated with the KTW model (183). Indeed, this theory proposes that bacteriophage will enter lysogenic life cycles either at low or high concentrations of their respective bacteria hosts, but will be lytic otherwise (183). The evolutionary benefit of this is apparent. If a host is present in high concentration, then the bacteriophage can integrate and replicate alongside the host and take advantage of the rapid replication rates. However, if the host is present in low concentration, then the bacteriophage can integrate to maintain itself while not providing stress on the host (183). The piggyback-the-winner theory has been expanded to mucosal surfaces (184) in conjunction with the bacteriophage-adhering mucosal model (181). This latter theory proposes the existence of a bacteria and bacteriophage gradient across mucosal surfaces (181). Towards the lumen, bacteria and bacteriophage concentrations are highest and operate under lysogenic-favored replication

(184). The deeper layers of the mucosa become bacteriophage rich and bacteria scarce, which shift bacteriophage towards the KTW dynamics, or lytic-cycle activation of the bacteriophage (184). Ultimately, high bacteriophage concentrations are noted nearest the epithelial surface where bacteria concentrations are lowest (184). The proximity of bacteriophage to the epithelium also provides a site for interactions with the human immune system.

BACTERIOPHAGE INTERACTIONS WITH THE HUMAN IMMUNE SYSTEM

Bacteriophage have been shown to act directly and/or indirectly with the mammalian immune system (185, 186). Caudovirales abundance is increased during inflammatory diseases, including IBD (156, 187); however, there is also evidence that bacteriophage directly cause inflammation. Examination of the immune system response in mouse models has shown that some bacteriophage can activate the immune system throught toll-like receptor (TLR)-9 mediated production of interferon (IFN)-gamma, (188) ultimately initiating both an adaptive T-cell response and exacerbating innate inflammation. Additionally, bacteriophage can influence the success of a fecal microbiota transplant (FMT). An FMT is the transfer of a donor's microbiota into a patient. An increased abundance of bacteriophage has been associated with FMT failure in IBD patients (188, 189).

Analysis of the virome in FMTs has also identified a stable core virome found in the human gastrointestinal tract (190). The core virome is a collection of bacteriophages that are shared across individuals (191, 192). *Crassphage* (75), for example, are part of this core virome in humans and represent members of one of the most abundant bacteriophage families (76, 77); they have not been associated with illness but were found to have a high degree of genetic diversity (193). While *Crassphage* was shown to infect *Bacteroides intestinalis* (77), *in-silico* analyses predict a broader host range (76). Microviridae were also suggested to comprise the core virome as they are

frequently found in humans (150, 151, 192, 194, 195) and animals (196) and have been shown to integrate as prophages in *Bacteroidetes* and *Prevotella* genera (197). Moreover, *Faecalibacterium prausnitzii* was indicated as a potential host of Microviridae since the presence of bacteriophage genes were identified in its genome (196). Additional investigation is needed to discover new viruses and evaluate their impact on human health.

EUKARYOTIC VIRUSES

Eukaryotic viruses are the other major component of the virome. The earlier studies on the soil virome utilized multiple-displacement amplification (198), which has an inherent bias to amplify circular DNA (199), and thus, the actual abundance of these viruses is unknown. Newer technologies have observed the virome of patients with IBD (156), HIV (38), non-polio acute flaccid paralysis (200), and hand-foot-mouth disease (201, 202). Previous studies in patients suffering from diarrhea (203, 204) have identified novel species of virus (203, 205–208), many of which belong to *Picobirnavirus* (203, 206, 207, 209, 210). *Picobirnavirus* is a double-stranded RNA virus that was thought to utilize mammals as hosts because of the high frequency of recovery from mammalian stools (211–214), yet they have not been successfully cultured in the laboratory (215). Although *Picobirnavirus* was found in 20% of cases with diarrheal illness in one study (210), its significance and function remain unclear. Recently it was suggested that invertebrates and even bacteria could be the hosts of *Picobirnavirus* because a conserved motif (ribosomal binding site) from prokaryotes was found in untranslated regions of the *Picobirnavirus* genome (215).

Additional eukaryotic virus families have been identified in the human gut, which include *Papillomaviridae*, *Polyomaviridae*, *Herpesviridae*, *Anelloviridae*, and *Circoviridae*. *Anelloviridae*, for instance, is diverse, comprising over 200 species, though they are not associated

with any diseases (216). *Anelloviridae* is found frequently in animals (217–219) and humans, namely the gastrointestinal tract (220, 221), respiratory tract (222), and cardiovascular system (223, 224). Although *Anelloviridae* has been reported to be elevated in disease states such as HIV (225), severe malnutrition (149), and malabsorption (226), these viruses have not been directly shown to cause disease. Further experimentation is needed to define the role of *Anelloviridae* in human health.

Similarly, in the human gut elevations in eukaryotic viruses such as *Mastadenovirus* and *Cytomegalovirus* (Herpesviridae), have both been observed in gestational diabetes (227). Mouse models utilizing murine *Cytomegalovirus* (mCMV) provide protection against infection from both *Yersinia pestis* and *Listeria monocytogenes* (228) by upregulating cytokine INF-gamma. This upregulation creates a higher elevated basal state of inflammation that wards against incoming infections and is not antigen-specific for the bacteria, though this enhanced immune response could also lead to more serious conditions such as autoimmunity (229–231) or cancer (232, 233). *Orthopoxvirus* was also found to be elevated in the meconium in gestational diabetes in humans (227) but decreased Proteobacteria abundance was observed in mouse models of *Orthopoxvirus* infection (234). Importantly, *Orthopoxvirus* produces soluble molecules that bind chemokines, cytokines and interferon to dampen host immune responses (235, 236). Other eukaryotic viruses such as *Norovirus*, have been shown to affect the immune system. *Norovirus* inoculations in germ-free mice, for instance, failed to elicit an immune response and restored the morphology of the intestinal tract that was affected by colitis (237). Inactivated *Rotavirus* could also reduce inflammation via activation of anti-inflammatory cytokines acting on toll-like receptors (238).

In summary, both bacteriophage and eukaryotic viruses have wide-reaching effects on the bacterial microbiota and the human host. Bacteriophage can directly infect and affect bacterial

microbiota, transfer gene amongst different bacterial populations, and alter host physiology. Bacteriophage can directly interact with the human immune system in a pro-inflammatory manner. Eukaryotic viruses can infect the human host and alter immune cell responses; however, additional studies are needed to determine the significance of these viruses within a microbial community and association with disease.

THE MICROBIOME OF VIRAL GASTROENTERITIS

Viral gastroenteritis refers to a gastrointestinal infection caused by a virus, and the symptoms include vomiting and watery diarrhea. In a prior study, the microbiome of pediatric patients with acute viral gastroenteritis (n=20, 15 *Norovirus*, 5 *Rotavirus*) was compared to healthy controls (n=20). Patients were stratified by mild versus severe disease based on the clinical presentation (239). Patients with severe disease had decreased Shannon diversity compared to both the mild patient and healthy control groups (239). Additionally, *Norovirus* infections did not appear to alter the microbiome as noted in other studies (123, 240, 241), nor did it contribute to an increase in inflammatory markers like lactoferrin (240). *Rotavirus*, however, caused a significant decrease in Shannon diversity with decreases in *Rikenellaceae*, *Porphyromonadaceae*, and *Alistipes* (239). *Parabacteroides* were found in equal proportions among cases with both severe and mild forms of viral gastroenteritis (239). Additionally, *Prevotellaceae*, *Staphylococcaceae*, and *Coriobacteriaceae*, specifically *Prevotella*, TM7, *Atopobium*, and *Staphylococcus*, were associated with abdominal pain (239). *Staphylococcus* has also been correlated in a previous study with abdominal pain in children (242). Convulsions were associated with decreased abundance of *Haemophilus* and *Faecalbacterium*, while viral gastroenteritis patients with complications had an increased abundance of *Campylobacteraceae*, *Neisseriaceae*, *Methylobacteriaceae*, *Sphingomonadaceae*, and *Enterobacteriaceae* (239). *Enterobacteriaceae*, specifically *E. coli*,

were the only taxa elevated in patients with *Norovirus* infections (240). Stool consistency has been correlated with *Norovirus* infection (243), which coincides with associations seen in stool patterns related to bacterial richness and diversity in bacterial gastroenteritis (123, 244). Secondary infections due to viruses are a common occurrence in respiratory tract infections (245) but remain wholly understudied in gastrointestinal illness. One gastrointestinal study of mixed infections involving bacteria and viruses in children identified that more severe disease resulted when only one infectious agent was present (246). Alterations in the microbiome due to mixed infections included reduced diversity in *Bacteroidetes* and increased richness in *Bifidobacteriaceae*, which was correlated with disease severity (246).

THE MICROBIOME OF BACTERIAL GASTROENTERITIS

Bacterial gastroenteritis is an infection of the gastrointestinal tract by a bacterial pathogen and can present typically with bloody diarrhea as a distinguishing feature relative to viral gastroenteritis. Bacterial pathogens such as *Salmonella* can directly cause diarrheal illness by exploiting inflammation to create a niche for colonizing the gastrointestinal tract (247), subsequently altering the microbiota. Alterations in the microbiota due to a pathogen have been observed in mouse models of *Citrobacteria rodentium* (26), and the intestinal microbiota is restored to a pre-infection state once the pathogen is cleared (22).

Enterotypes are a grouping of samples based on a dominant phylum (248). Patients with acute bacterial gastroenteritis have a shift in their microbiome to an *Escherichia-Shigella* enterotype (123). The dysbiosis from a bacterial pathogen affects three significant components of the microbiome. One, there is a decrease in short-chain fatty acid producers. Two, there is an increase in inflammation both due to a loss of small-chain fatty acid producers and anti-inflammatory bacteria. Three, a commensal bacterium can then bloom and continue the disease

process as observed with AIEC (26). The amount of dysbiosis that occurs in an illness can be correlated with the severity of disease (242).

The dysbiosis observed in bacterial gastroenteritis patients includes an increased abundance of Proteobacteria and a decrease in the Firmicutes:Bacteroides ratio (40, 123, 241). The dysbiosis that occurs is not pathogen-specific (40, 123, 241) and aside from increased abundance of Proteobacteria and decreases in Firmicutes and Bacteroides, there is little agreement among available bacterial gastroenteritis studies. For example, *Lacnospiracae*, a family of small-chain fatty acid producers (249), was reported to be increased in abundance in two reports (123, 241) and decreased in abundance in other reports (40). (250); however, additional studies are needed to determine the alterations that occur with *Lachnospiracae* in bacterial gastroenteritis.

Roseburia, another small-chain fatty acid producer, was reported to be in decreased abundance in one study (40), increased in abundance in another study (241) and not significantly different in abundance in a third study (123). *Faecalbacterium* can produce anti-inflammatory effects (251) and is decreased in abundance in some cases of gastroenteritis (40, 123). Decreased *Rikenellaceae* abundance has been associated with inflammation in IBD patients (252) and likewise was lower in abundance in gastroenteritis cases (40, 123).

Bilophila, a common commensal from the Proteobacteria phylum, was observed marginally increased in abundance in cases in one study (241), but these changes were not observed in two other studies of gastroenteritis (40, 123). *Bilophila* is bile-resistant and has been isolated from clinical specimens (253), which suggests it could be an opportunistic pathogen that arises during dysbiosis. Another opportunistic pathogen could be *Streptococcaceae,* which is a common commensal, but has been reported to be higher in abundance in gastroenteritis cases (40, 123). Increased abundance of *Streptococcaceae* has been associated with gut inflammation (254,

255). Another common commensal found throughout the gastrointestinal tract is *Veillonellaceae* (256). *Veillonellaceae* can hydrolyze bile-acids (257). Genera of *Veillonellaceae* ar higher in abundance in gastroenteritis cases (123). Additional studies are needed to confirm these findings. The metabolic profiles of gastroenteritis have been investigated using PICRUSt (123).

Six metabolic pathways were enriched within the microbiome of stool samples that also exhibited a higher abundance of Proteobacteria (123). These included cytochrome P450 related genes were enriched which are essential for drug metabolism. Bacterial associated genes were elevated and included bacterial invasion genes and lipopolysaccharide biosynthesis proteins (123). Liposaccharides constitute a significant component of the bacterial cell wall and are inflammatory if derived from Proteobacteria (258, 259) and anti-inflammatory if derived from *Bacteroidetes* (260). Structural differences in LPS between species impact the inflammatory response differently (260). Other immune response system associated genes were also impacted. The RIG-I-like receptor signalling pathway was elevated in microbiomes with increased abundance of Proteobacteria (123). This finding is unclear since bacteria have not been shown to activate the RIG-I pathway directly (261). Additionally, glycan metabolism pathways were enriched in gastroenteritis patients (123). Glycans are important for adhesion to mucosal surfaces (262) and are involved in adhesion to the mucosa. Conversely, flavonol biosynthesis genes were decreased (123). Flavonol blocks the adhesion of *E. coli* to surfaces (263) and directly alters the composition of the microbiome based on consumption (264). Additional analysis by the same group (123) identified that cases with the *Escherichia-Shigella* enterotype have enrichment in pathways related to bacterial invasion of epithelial cells, RIG-I-like receptor signaling, lipopolysaccharide biosynthesis proteins, and enrichment of proinflammatory pathways. Further studies should focus on an integrative approach integrating metatranscriptomics and metagenomics to elucidate both

the metabolically potential and activity of the microbiome, as was done with the recent human microbiome project with IBD (103). Future studies need to sample multiple time points during an acute infection to understand longitudinal changes in the microbiome.

CURRENT CHALLENGES AND GAPS

Current research into gastroenteritis (both viral and bacterial) has focused on studying either the bacterial component of the microbiome (40, 123, 241) or the viral component separately (206, 243). Studies of the microbiome that have performed network analysis to identify correlations between the virome and bacteriome (38, 156, 265) have provided great insight into the ecology of microbial communities and have highlighted new avenues to investigate disease pathogenesis. To date, there have been few sufficiently powered studies to analyse the virome and bacteriome together in acute bacterial gastroenteritis. Additionally, viral databases remain wholly incomplete, and additional isolation and characterization of viruses are therefore needed.

In order to address these knowledge gaps, this study was conducted with the following objectives:

1. Determine the organisms of the microbiome in acute bacterial gastroenteritis patients compared to non-infected controls that correlate with disease presentation.

Hypotheses: *The microbiome of gastrointestinal patients will be distinct from that of their healthy family member controls.*

2. Determine the organisms of the microbiome in acute bacterial gastroenteritis patients compared to their recovery state to identify changes over time.

Hypotheses: *The microbiome from recovered gastrointestinal patients will have profiles distinct from their infected microbiome.*

3. Characterize the functionality of bacteriophage isolated from intestinal viral communities.

Hypotheses: *The virome will have distinct functional profiles that are linked to the presence and abundance of Caudovirales.*

In all, this project will characterize the intestinal microbiomes of healthy controls and patients with acute bacterial gastroenteritis as well as a subset of the same patients after recovery. The network analysis will provide the most comprehensive picture of the microbiome of acute bacterial gastroenteritis to date. The matched cohort study of the cases and their follow-up state will directly assess for alterations in the microbiome. Additionally, bacteriophages will be isolated and characterized to add to existing databases and evaluate their ability to infect common enteric pathogens.

APPENDIX

Figure 1.1. Lytic/lysogenic conversion of resident bacteriophage Adapted from (183).

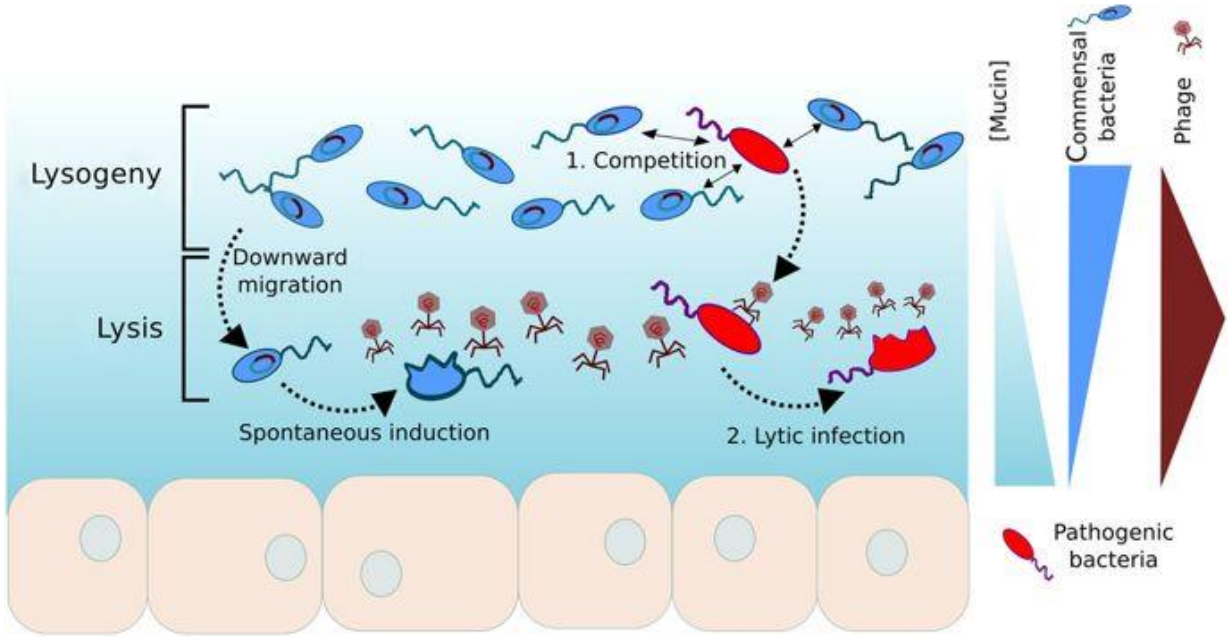
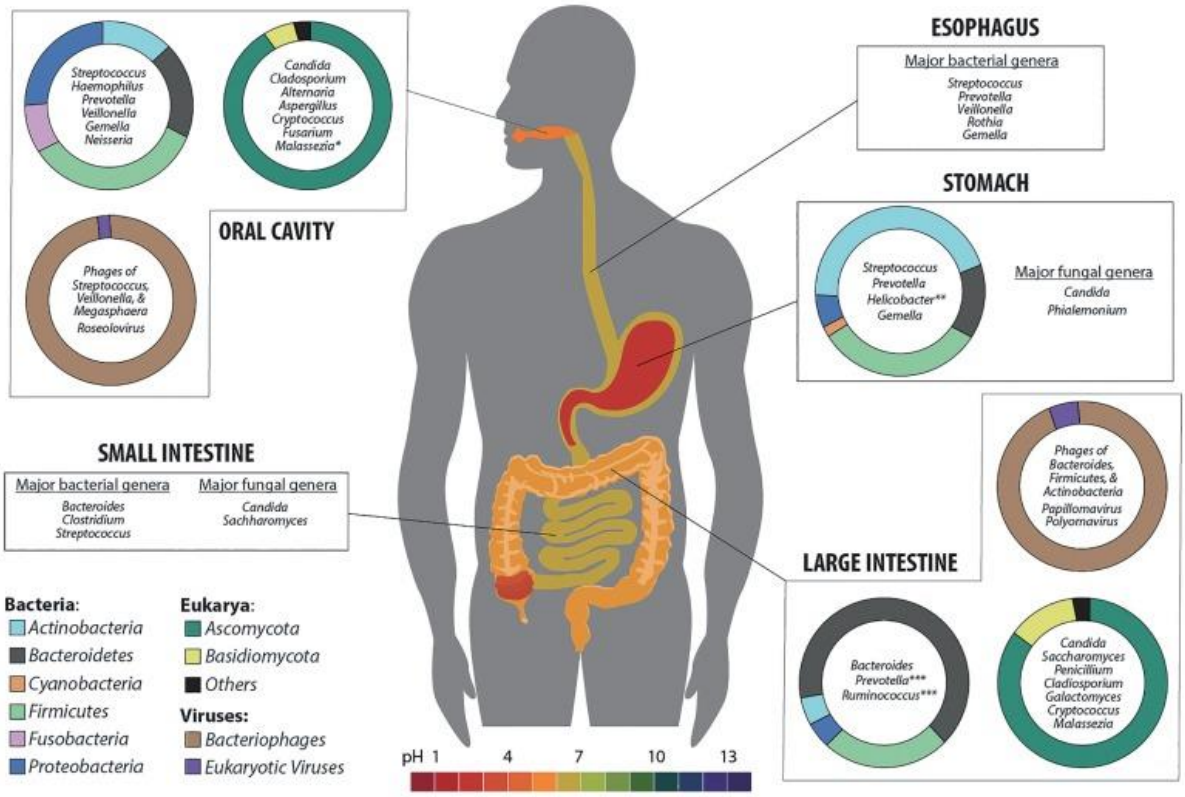


Figure 1.2. Microbiome composition of the gastrointestinal tract Adapted from (266).



REFERENCES

REFERENCES

1. **World Health Organization**. 2017. Diarrhoeal disease. Clin Med.

2. **Vos T**, **Allen C**, **Arora M**, **Barber RM**, **Bhutta ZA**, **Brown A**, **Carter A**, **Casey DC**, **Charlson FJ**, **Chen AZ**, **Coggeshall M**, **Cornaby L**, **Dandona L**, **Dicker DJ**,. 2016. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet **388**:1545–1602.

3. **UNICEF.** Diarrhoeal Disease: Current Status + Progress.

4. **Kotloff KL**. 2017. The Burden and Etiology of Diarrheal Illness in Developing Countries. Pediatr Clin North Am **64**:799–814.

5. **Hall AJ**, **Rosenthal M**, **Gregoricus N**, **Greene SA**, **Ferguson J**, **Henao OL**, **Vinjé J**, **Lopman BA**, **Parashar UD**, **Widdowson MA**. 2011. Incidence of acute gastroenteritis and role of norovirus, Georgia, USA, 2004-2005. Emerg Infect Dis **17**:1381–1388.

6. **Herikstad H**, **Yang S**, **Van Gilder TJ**, **Vugia D**, **Hadler J**, **Blake P**, **Deneen V**, **Shiferaw B**, **Angulo FJ**, **The FOODNET Working Group**. 2002. A population-based estimate of the burden of diarrhoeal illness in the United States: FoodNet, 1996–7. Epidemiol Infect **129**:9–17.

7. **Fletcher SM**, **Stark D**, **Ellis J**. 2011. Prevalence of gastrointestinal pathogens in Sub-Saharan Africa: systematic review and meta-analysis. J Public health Res **2**:e30.

8. **American Academy of Family Physicians., Hartman S**, **Brown E**, **Loomis E**, **Russell HA**. 2019. Gastroenteritis in Children. Am Fam Physician **99**:159–165.

9. **Rohner P**, **Pittet D**, **Pepey B**, **Nije-Kinge T**, **Auckenthaler R**. 1997. Etiological agents of infectious diarrhea: Implications for requests for microbial culture. J Clin Microbiol **35**:1427–1432.

10. **Chau ML**, **Hartantyo SHP**, **Yap M**, **Kang JSL**, **Aung KT**, **Gutiérrez RA**, **Ng LC**, **Tam CC**, **Barkham T**. 2016. Diarrheagenic pathogens in adults attending a hospital in Singapore. BMC Infect Dis **16**:1–9.

11. **Koplan JP**, **Jane Benfari Ferraro M**, **Fineberg H V.**, **Rosenberg ML**. 1980. VALUE OF STOOL CULTURES. Lancet.

12. **Braun T**, **Di Segni A**, **BenShoshan M**, **Asaf R**, **Squires JE**, **Farage Barhom S**, **Glick Saar E**, **Cesarkas K**, **Smollan G**, **Weiss B**, **Amit S**, **Keller N**, **Haberman Y**. 2017. Fecal microbial characterization of hospitalized patients with suspected infectious diarrhea shows significant dysbiosis. Sci Rep **7**:1088.

13. **Lee JY**, **Cho SY**, **Hwang HSH**, **Ryu JY**, **Lee J**, **Song I Do**, **Kim BJ**, **Kim JW**, **Chang SK**, **Choi CH**. 2017. Diagnostic yield of stool culture and predictive factors for positive

culture in patients with diarrheal illness. Med (United States) **96**:2–6.

14. **Murphy MS**. 1998. Guidelines for managing acute gastroenteritis based on a systematic review of published research. Arch Dis Child **79**:279–284.

15. **Centers for Disease Control and Prevention**. 2004. Diagnosis and Management of Foodborne Illnesses A Primer for Physicians and Other Health Care Professionals Centers for Disease Control and Prevention Epidemiology Program Office. MMWR.

16. **American Academy of Family Physicians., Hartman S**, **Brown E**, **Loomis E**, **Russell HA**. 2019. Gastroenteritis in Children. Am Fam Physician **99**:159–165.

17. **Cadwgan AM**, **Watson WA**, **Laing RBS**, **MacKenzie AR**, **Smith CC**, **Douglas JG**. 2000. Presenting clinical features and C-reactive protein in the prediction of a positive stool culture in patients with diarrhoea. J Infect.

18. **Kirk MD**, **Pires SM**, **Black RE**, **Caipo M**, **Crump JA**, **Devleesschauwer B**, **Döpfer D**, **Fazil A**, **Fischer-Walker CL**, **Hald T**, **Hall AJ**, **Keddy KH**, **Lake RJ**, **Lanata CF**, **Torgerson PR**, **Havelaar AH**, **Angulo FJ**. 2015. World Health Organization Estimates of the Global and Regional Disease Burden of 22 Foodborne Bacterial, Protozoal, and Viral Diseases, 2010: A Data Synthesis. PLoS Med **12**:1–21.

19. **Kotloff KL**, **Nataro JP**, **Blackwelder WC**, **Nasrin D**, **Farag TH**, **Panchalingam S**, **Wu Y**, **Sow SO**, **Sur D**, **Breiman RF**, **Faruque ASG**, **Zaidi AKM**, **Saha D**, **Alonso PL**, **Tamboura B**, **Sanogo D**, **Onwuchekwa U**, **Manna B**, **Ramamurthy T**, **Kanungo S**, **Ochieng JB**, **Omore R**, **Oundo JO**, **Hossain A**, **Das SK**, **Ahmed S**, **Qureshi S**, **Quadri F**, **Adegbola RA**, **Antonio M**, **Hossain MJ**, **Akinsola A**, **Mandomando I**, **Nhampossa T**, **Acácio S**, **Biswas K**, **O'Reilly CE**, **Mintz ED**, **Berkeley LY**, **Muhsen K**, **Sommerfelt H**, **Robins-Browne RM**, **Levine MM**. 2013. 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. Lancet **382**:209–222.

20. **Ijaz MK**, **Rubino JR**. 2012. Impact of infectious diseases on cognitive development in childhood and beyond: Potential mitigational role of hygiene. Open Infect Dis J.

21. **Bradshaw C**, **Zheng Y**, **Silver SA**, **Chertow GM**, **Long J**, **Anand S**. 2018. Acute Kidney Injury Due to Diarrheal Illness Requiring Hospitalization: Data from the National Inpatient Sample. J Gen Intern Med **33**:1520–1527.

22. **Lupp C**, **Robertson ML**, **Wickham ME**, **Sekirov I**, **Champion OL**, **Gaynor EC**, **Finlay BB**. 2007. Host-Mediated Inflammation Disrupts the Intestinal Microbiota and Promotes the Overgrowth of Enterobacteriaceae. Cell Host Microbe **2**:119–129.

23. **Schwille-Kiuntke J**, **Enck P**, **Zendler C**, **Krieg M**, **Polster A V.**, **Klosterhalfen S**, **Autenrieth IB**, **Zipfel S**, **Frick JS**. 2011. Postinfectious irritable bowel syndrome: Follow-up of a patient cohort of confirmed cases of bacterial infection with Salmonella or Campylobacter. Neurogastroenterol Motil.

24. **Thabane M**, **Simunovic M**, **Akhtar-Danesh N**, **Garg AX**, **Clark WF**, **Collins SM**,

Salvadori M, **Marshall JK**. 2010. An outbreak of acute bacterial gastroenteritis is associated with an increased incidence of irritable bowel syndrome in children. Am J Gastroenterol.

25. **Rodríguez LAG**, **Ruigómez A**, **Panés J**. 2006. Acute Gastroenteritis Is Followed by an Increased Risk of Inflammatory Bowel Disease. Gastroenterology **130**:1588–1594.

26. **Small CL**, **Xing L**, **McPhee JB**, **Law HT**, **Coombes BK**. 2016. Acute Infectious Gastroenteritis Potentiates a Crohn's Disease Pathobiont to Fuel Ongoing Inflammation in the Post-Infectious Period. PLoS Pathog **12**:1–20.

27. **Mohr JL**. 1952. Protozoa as Indicators of Pollution. Sci Mon.

28. **Lederberg J**, **McCray A**. 2001. 'Ome Sweet 'Omics-- A Genealogical Treasury of Words | The Scientist Magazine®. Sci.

29. **David LA**, **Maurice CF**, **Carmody RN**, **Gootenberg DB**, **Button JE**, **Wolfe BE**, **Ling A V.**, **Devlin AS**, **Varma Y**, **Fischbach MA**, **Biddinger SB**, **Dutton RJ**, **Turnbaugh PJ**. 2014. Diet rapidly and reproducibly alters the human gut microbiome. Nature.

30. **Francino MP**. 2016. Antibiotics and the human gut microbiome: Dysbioses and accumulation of resistances. Front Microbiol.

31. **Goodrich JK**, **Waters JL**, **Poole AC**, **Sutter JL**, **Koren O**, **Blekhman R**, **Beaumont M**, **Van Treuren W**, **Knight R**, **Bell JT**, **Spector TD**, **Clark AG**, **Ley RE**. 2014. Human genetics shape the gut microbiome. Cell **159**:789–799.

32. **Yatsunenko T**, **Rey FE**, **Manary MJ**, **Trehan I**, **Dominguez-Bello MG**, **Contreras M**, **Magris M**, **Hidalgo G**, **Baldassano RN**, **Anokhin AP**, **Heath AC**, **Warner B**, **Reeder J**, **Kuczynski J**, **Caporaso JG**, **Lozupone CA**, **Lauber C**, **Clemente JC**, **Knights D**, **Knight R**, **Gordon JI**. 2012. Human gut microbiome viewed across age and geography. Nature.

33. **Mariat D**, **Firmesse O**, **Levenez F**, **Guimarăes VD**, **Sokol H**, **Doré J**, **Corthier G**, **Furet JP**. 2009. The firmicutes/bacteroidetes ratio of the human microbiota changes with age. BMC Microbiol.

34. **Fredrik B**. 2015. Insights Into the Role of the Microbiome in Obesity and Type 2 Diabetes **38**:159–165.

35. **Turnbaugh PJ**, **Ley RE**, **Mahowald MA**, **Magrini V**, **Mardis ER**, **Gordon JI**. 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature.

36. **Plottel CS**, **Blaser MJ**. 2011. Microbiome and malignancy. Cell Host Microbe.

37. **Sheehan D**, **Moran C**, **Shanahan F**. 2015. The microbiota in inflammatory bowel disease. J Gastroenterol 495–507.

38. **Monaco CL**, **Gootenberg DB**, **Zhao G**, **Handley SA**, **Ghebremichael MS**, **Lim ES**, **Lankowski A**, **Baldridge MT**, **Wilen CB**, **Flagg M**, **Norman JM**, **Keller BC**, **Lu??vano JM**, **Wang D**, **Boum Y**, **Martin JN**, **Hunt PW**, **Bangsberg DR**, **Siedner MJ**, **Kwon DS**,

Virgin HW. 2016. Altered Virome and Bacterial Microbiome in Human Immunodeficiency Virus-Associated Acquired Immunodeficiency Syndrome. Cell Host Microbe **19**:311–322.

39. **Zhang X**, **Zhang D**, **Jia H**, **Feng Q**, **Wang D**, **Liang D**, **Wu X**, **Li J**, **Tang L**, **Li Y**, **Lan Z**, **Chen B**, **Li Y**, **Zhong H**, **Xie H**, **Jie Z**, **Chen W**, **Tang S**, **Xu X**, **Wang X**, **Cai X**, **Liu S**, **Xia Y**, **Li J**, **Qiao X**, **Al-Aama JY**, **Chen H**, **Wang L**, **Wu QJ**, **Zhang F**, **Zheng W**, **Li Y**, **Zhang M**, **Luo G**, **Xue W**, **Xiao L**, **Li J**, **Chen W**, **Xu X**, **Yin Y**, **Yang H**, **Wang J**, **Kristiansen K**, **Liu L**, **Li T**, **Huang Q**, **Li Y**, **Wang J**. 2015. The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. Nat Med.

40. **Singh P**, **Teal TK**, **Marsh TL**, **Tiedje JM**, **Mosci R**, **Jernigan K**, **Zell A**, **Newton DW**, **Salimnia H**, **Lephart P**, **Sundin D**, **Khalife W**, **Britton RA**, **Rudrik JT**, **Manning SD**. 2015. Intestinal microbial communities associated with acute enteric infections and disease recovery. Microbiome **3**:45.

41. **Pace NR**, **Stahl DA**, **Lane DJ**, **Olsen GJ**. 1986. The Analysis of Natural Microbial Populations by Ribosomal RNA Sequences.

42. **Marchesi JR**, **Ravel J**. 2015. The vocabulary of microbiome research: a proposal. Microbiome.

43. **Case RJ**, **Boucher Y**, **Dahllöf I**, **Holmström C**, **Doolittle WF**, **Kjelleberg S**. 2007. Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies. Appl Environ Microbiol.

44. **Hadziavdic K**, **Lekang K**, **Lanzen A**, **Jonassen I**, **Thompson EM**, **Troedsson C**. 2014. Characterization of the 18s rRNA gene for designing universal eukaryote specific primers. PLoS One.

45. **Schoch CL**, **Seifert KA**, **Huhndorf S**, **Robert V**, **Spouge JL**, **Levesque CA**, **Chen W**, **Bolchacova E**, **Voigt K**, **Crous PW**, **Miller AN**, **Wingfield MJ**, **Aime MC**, **An KD**, **Bai FY**, **Barreto RW**, **Begerow D**, **Bergeron MJ**, **Blackwell M**, **Boekhout T**, **Bogale M**, **Boonyuen N**, **Burgaz AR**, **Buyck B**, **Cai L**, **Cai Q**, **Cardinali G**, **Chaverri P**, **Coppins BJ**, **Crespo A**, **Cubas P**, **Cummings C**, **Damm U**, **de Beer ZW**, **de Hoog GS**, **Del-Prado R**, **Dentinger B**, **Diéguez-Uribeondo J**, **Divakar PK**, **Douglas B**, **Dueñas M**, **Duong TA**, **Eberhardt U**, **Edwards JE**, **Elshahed MS**, **Fliegerova K**, **Furtado M**, **García MA**, **Ge ZW**, **Griffith GW**, **Griffiths K**, **Groenewald JZ**, **Groenewald M**, **Grube M**, **Gryzenhout M**, **Guo LD**, **Hagen F**, **Hambleton S**, **Hamelin RC**, **Hansen K**, **Harrold P**, **Heller G**, **Herrera C**, **Hirayama K**, **Hirooka Y**, **Ho HM**, **Hoffmann K**, **Hofstetter V**, **Högnabba F**, **Hollingsworth PM**, **Hong SB**, **Hosaka K**, **Houbraken J**, **Hughes K**, **Huhtinen S**, **Hyde KD**, **James T**, **Johnson EM**, **Johnson JE**, **Johnston PR**, **Jones EBG**, **Kelly LJ**, **Kirk PM**, **Knapp DG**, **Kõljalg U**, **Kovács GM**, **Kurtzman CP**, **Landvik S**, **Leavitt SD**, **Liggenstoffer AS**, **Liimatainen K**, **Lombard L**, **Luangsa-ard JJ**, **Lumbsch HT**, **Maganti H**, **Maharachchikumbura SSN**, **Martin MP**, **May TW**, **McTaggart AR**, **Methven AS**, **Meyer W**, **Moncalvo JM**, **Mongkolsamrit S**, **Nagy LG**, **Nilsson RH**, **Niskanen T**, **Nyilasi I**, **Okada G**, **Okane I**, **Olariaga I**, **Otte J**, **Papp T**, **Park D**, **Petkovits T**, **Pino-Bodas R**, **Quaedvlieg W**, **Raja HA**, **Redecker D**, **Rintoul TL**, **Ruibal C**, **Sarmiento-Ramírez JM**, **Schmitt I**, **Schüßler A**, **Shearer C**, **Sotome K**, **Stefani FOP**,

Stenroos S, **Stielow B**, **Stockinger H**, **Suetrong S**, **Suh SO**, **Sung GH**, **Suzuki M**, **Tanaka K**, **Tedersoo L**, **Telleria MT**, **Tretter E**, **Untereiner WA**, **Urbina H**, **Vágvölgyi C**, **Vialle A**, **Vu TD**, **Walther G**, **Wang QM**, **Wang Y**, **Weir BS**, **Weiß M**, **White MM**, **Xu J**, **Yahr R**, **Yang ZL**, **Yurkov A**, **Zamora JC**, **Zhang N**, **Zhuang WY**, **Schindel D**. 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proc Natl Acad Sci U S A.

46. **Callahan BJ**, **McMurdie PJ**, **Holmes SP**. 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J.

47. **Caporaso JG**, **Kuczynski J**, **Stombaugh J**, **Bittinger K**, **Bushman FD**, **Costello EK**, **Fierer N**, **Pẽa AG**, **Goodrich JK**, **Gordon JI**, **Huttley GA**, **Kelley ST**, **Knights D**, **Koenig JE**, **Ley RE**, **Lozupone CA**, **McDonald D**, **Muegge BD**, **Pirrung M**, **Reeder J**, **Sevinsky JR**, **Turnbaugh PJ**, **Walters WA**, **Widmann J**, **Yatsunenko T**, **Zaneveld J**, **Knight R**. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods.

48. **Schloss PD**, **Westcott SL**, **Ryabin T**, **Hall JR**, **Hartmann M**, **Hollister EB**, **Lesniewski RA**, **Oakley BB**, **Parks DH**, **Robinson CJ**, **Sahl JW**, **Stres B**, **Thallinger GG**, **Van Horn DJ**, **Weber CF**. 2009. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol.

49. **Maidak BL**, **Olsen GJ**, **Larsen N**, **Overbeek R**, **McCaughey MJ**, **Woese CR**. 1996. The Ribosomal Database Project (RDP). Nucleic Acids Res.

50. **Quast C**, **Pruesse E**, **Yilmaz P**, **Gerken J**, **Schweer T**, **Yarza P**, **Peplies J**, **Glöckner FO**. 2013. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Res.

51. **DeSantis TZ**, **Hugenholtz P**, **Larsen N**, **Rojas M**, **Brodie EL**, **Keller K**, **Huber T**, **Dalevi D**, **Hu P**, **Andersen GL**. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol.

52. **Eloe-Fadrosh EA**, **Ivanova NN**, **Woyke T**, **Kyrpides NC**. 2016. Metagenomics uncovers gaps in amplicon-based detection of microbial diversity. Nat Microbiol.

53. **Riesenfeld CS**, **Schloss PD**, **Handelsman J**. 2004. Metagenomics: Genomic Analysis of Microbial Communities. Annu Rev Genet.

54. **Martin M**. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal.

55. **Bolger a. M**, **Lohse M**, **Usadel B**. 2014. Trimmomatic: A flexible read trimming tool for Illumina NGS data. Bioinformatics **30**:2114–2120.

56. **Gordon A**, **Hannon GJ**. 2010. Fastx-toolkit. FASTQ/A short-reads pre-processing tools. Unpubl http//hannonlab cshl edu/fastx_ toolkit.

57. **Bushnell B**, **Rood J**, **Singer E**. 2017. BBTools Software Package. PLoS One.

58. **Andrews S**. 2010. FastQC: A quality control tool for high throughput sequence data. Http://WwwBioinformaticsBabrahamAcUk/Projects/Fastqc/.

59. **Ewels P**, **Magnusson M**, **Lundin S**, **Käller M**. 2016. MultiQC: Summarize analysis results for multiple tools and samples in a single report. Bioinformatics.

60. **Langmead B**, **Salzberg SL**, **Langmead**. 2013. Bowtie2. Nat Methods **9**:357–359.

61. **Li H**. 2010. Aligning new-sequencing reads by BWA BWA : Burrows-Wheeler Aligner. PPT.

62. **Wingett SW**, **Andrews S**. 2018. FastQ Screen: A tool for multi-genome mapping and quality control. F1000Research.

63. **Brown CT**, **Howe A**, **Zhang Q**, **Pyrkosz AB**, **Brom TH**. 2012. A single pass approach to reducing sampling variation, removing errors, and scaling {\em de novo} assembly of shotgun sequences. Pnas.

64. **Crusoe MR**, **Alameldin HF**, **Awad S**, **Boucher E**, **Caldwell A**, **Cartwright R**, **Charbonneau A**, **Constantinides B**, **Edvenson G**, **Fay S**, **Fenton J**, **Fenzl T**, **Fish J**, **Garcia-Gutierrez L**, **Garland P**, **Gluck J**, **González I**, **Guermond S**, **Guo J**, **Gupta A**, **Herr JR**, **Howe A**, **Hyer A**, **Härpfer A**, **Irber L**, **Kidd R**, **Lin D**, **Lippi J**, **Mansour T**, **McA'Nulty P**, **McDonald E**, **Mizzi J**, **Murray KD**, **Nahum JR**, **Nanlohy K**, **Nederbragt AJ**, **Ortiz-Zuazaga H**, **Ory J**, **Pell J**, **Pepe-Ranney C**, **Russ ZN**, **Schwarz E**, **Scott C**, **Seaman J**, **Sievert S**, **Simpson J**, **Skennerton CT**, **Spencer J**, **Srinivasan R**, **Standage D**, **Stapleton JA**, **Steinman SR**, **Stein J**, **Taylor B**, **Trimble W**, **Wiencko HL**, **Wright M**, **Wyss B**, **Zhang Q**, **Zyme E**, **Brown CT**. 2015. The khmer software package: enabling efficient nucleotide sequence analysis. F1000Research **4**:900.

65. **Oulas A**, **Pavloudi C**, **Polymenakou P**, **Pavlopoulos GA**, **Papanikolaou N**, **Kotoulas G**, **Arvanitidis C**, **Iliopoulos I**. 2015. Metagenomics: Tools and insights for analyzing next-generation sequencing data derived from biodiversity studies. Bioinform Biol Insights.

66. **Greninger AL**, **Messacar K**, **Dunnebacke T**, **Naccache SN**, **Federman S**, **Bouquet J**, **Mirsky D**, **Nomura Y**, **Yagi S**, **Glaser C**, **Vollmer M**, **Press CA**, **Kleinschmidt-DeMasters BK**, **Dominguez SR**, **Chiu CY**. 2015. Clinical metagenomic identification of Balamuthia mandrillaris encephalitis and assembly of the draft genome: The continuing case for reference genome sequencing. Genome Med.

67. **Berg MG**, **Lee D**, **Coller K**, **Frankel M**, **Aronsohn A**, **Cheng K**, **Forberg K**, **Marcinkus M**, **Naccache SN**, **Dawson G**, **Brennan C**, **Jensen DM**, **Hackett J**, **Chiu CY**. 2015. Discovery of a Novel Human Pegivirus in Blood Associated with Hepatitis C Virus Co-Infection. PLoS Pathog.

68. **Nagarajan N**, **Pop M**. 2013. Sequence assembly demystified. Nat Rev Genet.

69. **Zerbino DR**. 2010. Using the Velvet de novo assembler for short-read sequencing technologies. Curr Protoc Bioinforma.

70. **Bankevich A**, **Nurk S**, **Antipov D**, **Gurevich AA**, **Dvorkin M**, **Kulikov AS**, **Lesin VM**,

Nikolenko SI, **Pham S**, **Prjibelski AD**, **Pyshkin A V.**, **Sirotkin A V.**, **Vyahhi N**, **Tesler G**, **Alekseyev MA**, **Pevzner PA**. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol **19**:455–477.

71. **Namiki T**, **Hachiya T**, **Tanaka H**, **Sakakibara Y**. 2012. MetaVelvet: An extension of Velvet assembler to de novo metagenome assembly from short sequence reads. Nucleic Acids Res **40**.

72. **Li D**, **Liu CM**, **Luo R**, **Sadakane K**, **Lam TW**. 2014. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics **31**:1674–1676.

73. **Peng Y**, **Leung HCM**, **Yiu SM**, **Chin FYL**. 2012. IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics **28**:1420–1428.

74. **Nurk S**, **Meleshko D**, **Korobeynikov A**, **Pevzner PA**. 2017. metaSPAdes: a new versatile metagenomic assembler (Supplementary Material). Genome Res **27**:824–834.

75. **Dutilh BE**, **Cassman N**, **McNair K**, **Sanchez SE**, **Silva GGZ**, **Boling L**, **Barr JJ**, **Speth DR**, **Seguritan V**, **Aziz RK**, **Felts B**, **Dinsdale EA**, **Mokili JL**, **Edwards RA**. 2014. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. Nat Commun.

76. **Yutin N**, **Makarova KS**, **Gussow AB**, **Krupovic M**, **Segall A**, **Edwards RA**, **Koonin E V.** 2018. Discovery of an expansive bacteriophage family that includes the most abundant viruses from the human gut. Nat Microbiol **3**:38–46.

77. **Shkoporov AN**, **Khokhlova E V.**, **Fitzgerald CB**, **Stockdale SR**, **Draper LA**, **Ross RP**, **Hill C**. 2018. ΦCrAss001 represents the most abundant bacteriophage family in the human gut and infects Bacteroides intestinalis. Nat Commun **9**:1–8.

78. **Wilkins LGE**, **Ettinger CL**, **Jospin G**, **Eisen JA**. 2019. Metagenome-assembled genomes provide new insight into the microbial diversity of two thermal pools in Kamchatka, Russia. Sci Rep.

79. **Stewart RD**, **Auffret MD**, **Warr A**, **Wiser AH**, **Press MO**, **Langford KW**, **Liachko I**, **Snelling TJ**, **Dewhurst RJ**, **Walker AW**, **Roehe R**, **Watson M**. 2018. Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. Nat Commun **9**:1–11.

80. **Parks DH**, **Imelfort M**, **Skennerton CT**, **Hugenholtz P**, **Tyson GW**. 2015. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res.

81. **Wu YW**, **Simmons BA**, **Singer SW**. 2016. MaxBin 2.0: An automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics.

82. **Kang DD**, **Li F**, **Kirton E**, **Thomas A**, **Egan R**, **An H**, **Wang Z**. 2019. MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ.

83. **Imelfort M**, **Parks D**, **Woodcroft BJ**, **Dennis P**, **Hugenholtz P**, **Tyson GW**. 2014. GroopM: An automated tool for the recovery of population genomes from related metagenomes. PeerJ.

84. **Langille MGI**, **Zaneveld J**, **Caporaso JG**, **McDonald D**, **Knights D**, **Reyes JA**, **Clemente JC**, **Burkepile DE**, **Vega Thurber RL**, **Knight R**, **Beiko RG**, **Huttenhower C**. 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat Biotechnol.

85. **Handelsman J**, **Tiedje J**, **Alvarez-Cohen L**, **Ashburner M**, **Cann IKO**, **DeLong EF**, **Doolittle WF**, **Fraser-Liggett CM**, **Godzik A**, **Gordon JI**, **Riley M**, **Schmid M**. 2007. Revealing the Secrets of Our Microbial PlanetNational Academies Press.

86. **Moran MA**, **Satinsky B**, **Gifford SM**, **Luo H**, **Rivers A**, **Chan LK**, **Meng J**, **Durham BP**, **Shen C**, **Varaljay VA**, **Smith CB**, **Yager PL**, **Hopkinson BM**. 2013. Sizing up metatranscriptomics. ISME J.

87. **Altschul SF**, **Gish W**, **Miller W**, **Myers EW**, **Lipman DJ**. 1990. Basic local alignment search tool. J Mol Biol **215**:403–10.

88. **Benson DA**, **Cavanaugh M**, **Clark K**, **Karsch-Mizrachi I**, **Lipman DJ**, **Ostell J**, **Sayers EW**. 2013. GenBank. Nucleic Acids Res.

89. **Lindgreen S**, **Adair KL**, **Gardner PP**. 2016. An evaluation of the accuracy and speed of metagenome analysis tools. Sci Rep.

90. **Wood DE**, **Salzberg SL**. 2014. Kraken: Ultrafast metagenomic sequence classification using exact alignments. Genome Biol **15**.

91. **Ounit R**, **Wanamaker S**, **Close TJ**, **Lonardi S**. 2015. CLARK: fast and accurate classification of metagenomic and genomic sequences using discriminative k-mers. BMC Genomics.

92. **Buchfink B**, **Xie C**, **Huson DH**. 2014. Fast and sensitive protein alignment using DIAMOND. Nat Methods.

93. **Menzel P**, **Ng KL**, **Krogh A**. 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun **7**:1–9.

94. **Segata N**, **Waldron L**, **Ballarini A**, **Narasimhan V**, **Jousson O**, **Huttenhower C**. 2012. Metagenomic microbial community profiling using unique clade-specific marker genes. Nat Methods.

95. **Broder AZ**. 1997. On the resemblance and containment of documentsProceedings of the International Conference on Compression and Complexity of Sequences.

96. **Ondov BD**, **Treangen TJ**, **Melsted P**, **Mallonee AB**, **Bergman NH**, **Koren S**, **Phillippy AM**. 2016. Mash: Fast genome and metagenome distance estimation using MinHash. Genome Biol.

97. **Titus Brown C**, **Irber L**. 2016. sourmash: a library for MinHash sketching of DNA. J Open Source Softw.

98. **Turnbaugh PJ**, **Ley RE**, **Hamady M**, **Fraser-Liggett CM**, **Knight R**, **Gordon JI**. 2007. The Human Microbiome Project. Nature **449**:804–810.

99. **Rajilić-Stojanović M**, **de Vos WM**. 2014. The first 1000 cultured species of the human gastrointestinal microbiota. FEMS Microbiol Rev **38**:996–1047.

100. **Vaissi S**, **Sharifi M**, **Hernandez A**, **Nikpey S**, **Taran M**. 2019. Skin bacterial microflora of two closely related mountain newts (Salamandridae) – the Yellow-spotted mountain newt Neurergus derjugini and the Kaiser's mountain newt Neurergus kaiseri – in the wild and in a breeding facility highlight new conservation per. Int Zoo Yearb 1–11.

101. **Consortium THMP**. 2013. Structure, Function and Diversity of the Healthy Human Microbiome. Nature **486**:207–214.

102. **Ganji L**, **Alebouyeh M**, **Shirazi MH**, **Eshraghi SS**, **Mirshafiey A**, **Ebrahimi Daryani N**, **Zali MR**. 2016. Dysbiosis of fecal microbiota and high frequency of Citrobacter, Klebsiella spp., and Actinomycetes in patients with irritable bowel syndrome and gastroenteritis. Gastroenterol Hepatol from bed to bench **9**:325–330.

103. **Lloyd-Price J**, **Arze C**, **Ananthakrishnan AN**, **Schirmer M**, **Avila-Pacheco J**, **Poon TW**, **Andrews E**, **Ajami NJ**, **Bonham KS**, **Brislawn CJ**, **Casero D**, **Courtney H**, **Gonzalez A**, **Graeber TG**, **Hall AB**, **Lake K**, **Landers CJ**, **Mallick H**, **Plichta DR**, **Prasad M**, **Rahnavard G**, **Sauk J**, **Shungin D**, **Vázquez-Baeza Y**, **White RA**, **Braun J**, **Denson LA**, **Jansson JK**, **Knight R**, **Kugathasan S**, **McGovern DPB**, **Petrosino JF**, **Stappenbeck TS**, **Winter HS**, **Clish CB**, **Franzosa EA**, **Vlamakis H**, **Xavier RJ**, **Huttenhower C**. 2019. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature **569**:655–662.

104. **Chen J**, **Chia N**, **Kalari KR**, **Yao JZ**, **Novotna M**, **Soldan MMP**, **Luckey DH**, **Marietta E V.**, **Jeraldo PR**, **Chen X**, **Weinshenker BG**, **Rodriguez M**, **Kantarci OH**, **Nelson H**, **Murray JA**, **Mangalam AK**. 2016. Multiple sclerosis patients have a distinct gut microbiota compared to healthy controls. Sci Rep.

105. **Wang Z**, **Wang Q**, **Zhao J**, **Gong L**, **Zhang Y**, **Wang X**, **Yuan Z**. 2019. Altered diversity and composition of the gut microbiome in patients with cervical cancer. AMB Express **9**.

106. **Almonacid DE**, **Kraal L**, **Ossandon FJ**, **Budovskaya Y V**, **Cardenas JP**, **Bik EM**, **Goddard AD**, **Richman J**, **Apte ZS**. 2017. 16S rRNA gene sequencing and healthy reference ranges for 28 clinically relevant microbial taxa from the human gut microbiome. PLoS One **12**:e0176555.

107. **Ridaura VK**, **Faith JJ**, **Rey FE**, **Cheng J**, **Duncan AE**, **Kau AL**, **Griffin NW**, **Lombard V**, **Henrissat B**, **Bain JR**, **Muehlbauer MJ**, **Ilkayeva O**, **Semenkovich CF**, **Funai K**, **Hayashi DK**, **Lyle BJ**, **Martini MC**, **Ursell LK**, **Clemente JC**, **Van Treuren W**, **Walters WA**, **Knight R**, **Newgard CB**, **Heath AC**, **Gordon JI**. 2013. Gut microbiota from twins discordant for obesity modulate metabolism in mice. Science (80-).

108. **Tilg H**, **Kaser A**. 2011. Gut microbiome, obesity, and metabolic dysfunction. J Clin Invest.

109. **Arumugam M**, **Raes J**, **Pelletier E**, **Paslier D Le**, **Batto J**, **Bertalan M**, **Borruel N**, **Casellas F**, **Costea PI**, **Hildebrand F**, **Manimozhiyan A**, **Bäckhed F**, **Blaser MJ**, **Bushman FD**, **De Vos WM**, **Ehrlich SD**, **Fraser CM**, **Hattori M**, **Huttenhower C**, **Jeffery IB**, **Knights D**, **Lewis JD**, **Ley RE**, **Ochman H**, **O'Toole PW**, **Quince C**, **Relman DA**, **Shanahan F**, **Sunagawa S**, **Wang J**, **Weinstock GM**, **Wu GD**, **Zeller G**, **Zhao L**, **Raes J**, **Knight R**, **Bork P**, **Gorvitovskaia A**, **Holmes SP**, **Huse SM**. 2013. Enterotypes in the landscape of gut microbial community composition. Nature **3**:1–12.

110. **Koren O**, **Knights D**, **Gonzalez A**, **Waldron L**, **Segata N**, **Knight R**, **Huttenhower C**, **Ley RE**. 2013. A Guide to Enterotypes across the Human Body: Meta-Analysis of Microbial Community Structures in Human Microbiome Datasets. PLoS Comput Biol.

111. **Gorvitovskaia A**, **Holmes SP**, **Huse SM**. 2016. Interpreting prevotella and bacteroides as biomarkers of diet and lifestyle. Microbiome.

112. **Kovatcheva-Datchary P**, **Nilsson A**, **Akrami R**, **Lee YS**, **De Vadder F**, **Arora T**, **Hallen A**, **Martens E**, **Björck I**, **Bäckhed F**. 2015. Dietary Fiber-Induced Improvement in Glucose Metabolism Is Associated with Increased Abundance of Prevotella. Cell Metab.

113. **Ley RE**. 2016. Gut microbiota in 2015: Prevotella in the gut: Choose carefully. Nat Rev Gastroenterol Hepatol.

114. **Gupta VK**, **Chaudhari NM**, **Iskepalli S**, **Dutta C**. 2015. Divergences in gene repertoire among the reference Prevotella genomes derived from distinct body sites of human. BMC Genomics.

115. **Gao Z**, **Yin J**, **Zhang J**, **Ward RE**, **Martin RJ**, **Lefevre M**, **Cefalu WT**, **Ye J**. 2009. Butyrate improves insulin sensitivity and increases energy expenditure in mice. Diabetes.

116. **Lin H V.**, **Frassetto A**, **Kowalik EJ**, **Nawrocki AR**, **Lu MM**, **Kosinski JR**, **Hubert JA**, **Szeto D**, **Yao X**, **Forrest G**, **Marsh DJ**. 2012. Butyrate and propionate protect against diet-induced obesity and regulate gut hormones via free fatty acid receptor 3-independent mechanisms. PLoS One.

117. **Duncan SH**, **Holtrop G**, **Lobley GE**, **Calder AG**, **Stewart CS**, **Flint HJ**. 2004. Contribution of acetate to butyrate formation by human faecal bacteria. Br J Nutr.

118. **Le Chatelier E**, **Nielsen T**, **Qin J**, **Prifti E**, **Hildebrand F**, **Falony G**, **Almeida M**, **Arumugam M**, **Batto JM**, **Kennedy S**, **Leonard P**, **Li J**, **Burgdorf K**, **Grarup N**, **Jørgensen T**, **Brandslund I**, **Nielsen HB**, **Juncker AS**, **Bertalan M**, **Levenez F**, **Pons N**, **Rasmussen S**, **Sunagawa S**, **Tap J**, **Tims S**, **Zoetendal EG**, **Brunak S**, **Clément K**, **Doré J**, **Kleerebezem M**, **Kristiansen K**, **Renault P**, **Sicheritz-Ponten T**, **De Vos WM**, **Zucker JD**, **Raes J**, **Hansen T**, **Bork P**, **Wang J**, **Ehrlich SD**, **Pedersen O**, **Guedon E**, **Delorme C**, **Layec S**, **Khaci G**, **Van De Guchte M**, **Vandemeulebrouck G**, **Jamet A**, **Dervyn R**, **Sanchez N**, **Maguin E**, **Haimet F**, **Winogradski Y**, **Cultrone A**, **Leclerc M**, **Juste C**, **Blottière H**, **Pelletier E**, **Lepaslier D**, **Artiguenave F**, **Bruls T**, **Weissenbach J**, **Turner K**, **Parkhill J**, **Antolin M**, **Manichanh C**, **Casellas F**, **Boruel N**, **Varela E**, **Torrejon A**,

Guarner F, **Denariaz G**, **Derrien M**, **Van Hylckama Vlieg JET**, **Veiga P**, **Oozeer R**, **Knol J**, **Rescigno M**, **Brechot C**, **M'Rini C**, **Mérieux A**, **Yamada T**. 2013. Richness of human gut microbiome correlates with metabolic markers. Nature.

119. **Hansen EE**, **Lozupone CA**, **Rey FE**, **Wu M**, **Guruge JL**, **Narra A**, **Goodfellow J**, **Zaneveld JR**, **McDonald DT**, **Goodrich JA**, **Heath AC**, **Knight R**, **Gordon JI**. 2011. Pan-genome of the dominant human gut-associated archaeon, Methanobrevibacter smithii, studied in twins. Proc Natl Acad Sci U S A.

120. **Goodrich JK**, **Waters JL**, **Poole AC**, **Sutter JL**, **Koren O**, **Blekhman R**, **Beaumont M**, **Van Treuren W**, **Knight R**, **Bell JT**, **Spector TD**, **Clark AG**, **Ley RE**. 2014. Human genetics shape the gut microbiome. Cell **159**:789–799.

121. **Goodrich JK**, **Davenport ER**, **Beaumont M**, **Jackson MA**, **Knight R**, **Ober C**, **Spector TD**, **Bell JT**, **Clark AG**, **Ley RE**. 2016. Genetic Determinants of the Gut Microbiome in UK Twins. Cell Host Microbe.

122. **Nkamga VD**, **Henrissat B**, **Drancourt M**. 2017. Archaea: Essential inhabitants of the human digestive microbiota. Hum Microbiome J.

123. **Castaño-Rodríguez N**, **Underwood AP**, **Merif J**, **Riordan SM**, **Rawlinson WD**, **Mitchell HM**, **Kaakoush NO**. 2018. Gut microbiome analysis identifies potential etiological factors in acute gastroenteritis. Infect Immun **86**:1–13.

124. **Gérard P**. 2013. Metabolism of cholesterol and bile acids by the gut microbiota. Pathogens.

125. **Van Herreweghen F**, **De Paepe K**, **Roume H**, **Kerckhof FM**, **Van de Wiele T**. 2018. Mucin degradation niche as a driver of microbiome composition and Akkermansia muciniphila abundance in a dynamic gut model is donor independent. FEMS Microbiol Ecol.

126. **Desai MS**, **Seekatz AM**, **Koropatkin NM**, **Kamada N**, **Hickey CA**, **Wolter M**, **Pudlo NA**, **Kitamoto S**, **Terrapon N**, **Muller A**, **Young VB**, **Henrissat B**, **Wilmes P**, **Stappenbeck TS**, **N????ez G**, **Martens EC**. 2016. A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. Cell **167**:1339-1353.e21.

127. **O'Callaghan A**, **van Sinderen D**. 2016. Bifidobacteria and their role as members of the human gut microbiota. Front Microbiol.

128. **De Vuyst L**, **Leroy F**. 2011. Cross-feeding between bifidobacteria and butyrate-producing colon bacteria explains bifdobacterial competitiveness, butyrate production, and gas production. Int J Food Microbiol.

129. **Göker M**, **Gronow S**, **Zeytun A**, **Nolan M**, **Lucas S**, **Lapidus A**, **Hammon N**, **Deshpande S**, **Cheng JF**, **Pitluck S**, **Liolios K**, **Pagani I**, **Ivanova N**, **Mavromatis K**, **Ovchinikova G**, **Pati A**, **Tapia R**, **Han C**, **Goodwin L**, **Chen A**, **Palaniappan K**, **Land M**, **Hauser L**, **Jeffries CD**, **Brambilla EM**, **Rohde M**, **Detter JC**, **Woyke T**, **Bristow J**, **Markowitz V**, **Hugenholtz P**, **Eisen JA**, **Kyrpides NC**, **Klenk HP**. 2011. Complete genome sequence of

odoribacter splanchnicus type strain (1651/6 T). Stand Genomic Sci.

130. **Duncan SH**, **Hold GL**, **Barcenilla A**, **Stewart CS**, **Flint HJ**. 2002. Roseburia intestinalis sp. nov., a novel saccharolytic, butyrate-producing bacterium from human faeces. Int J Syst Evol Microbiol.

131. **Segain JP**, **Galmiche JP**, **Raingeard De La Blétière D**, **Bourreille A**, **Leray V**, **Gervois N**, **Rosales C**, **Ferrier L**, **Bonnet C**, **Blottière HM**. 2000. Butyrate inhibits inflammatory responses through NFκB inhibition: Implications for Crohn's disease. Gut.

132. **Quévrain E**, **Maubert MA**, **Michon C**, **Chain F**, **Marquant R**, **Tailhades J**, **Miquel S**, **Carlier L**, **Bermúdez-Humarán LG**, **Pigneur B**, **Lequin O**, **Kharrat P**, **Thomas G**, **Rainteau D**, **Aubry C**, **Breyner N**, **Afonso C**, **Lavielle S**, **Grill JP**, **Chassaing G**, **Chatel JM**, **Trugnan G**, **Xavier R**, **Langella P**, **Sokol H**, **Seksik P**. 2016. Identification of an anti-inflammatory protein from Faecalibacterium prausnitzii, a commensal bacterium deficient in Crohn's disease. Gut.

133. **Ness IF**, **Diep DB**, **Ike Y**. 2014. Enterococcal Bacteriocins and Antimicrobial Proteins that Contribute to Niche ControlEnterococci: From Commensals to Leading Causes of Drug Resistant Infection.

134. **Shin NR**, **Whon TW**, **Bae JW**. 2015. Proteobacteria: Microbial signature of dysbiosis in gut microbiota. Trends Biotechnol.

135. **Martin HM**, **Campbell BJ**, **Hart CA**, **Mpofu C**, **Nayar M**, **Singh R**, **Englyst H**, **Williams HF**, **Rhodes JM**. 2004. Enhanced Escherichia coli adherence and invasion in Crohn's disease and colon cancer. Gastroenterology.

136. **Darfeuille-Michaud A**, **Boudeau J**, **Bulois P**, **Neut C**, **Glasser AL**, **Barnich N**, **Bringer MA**, **Swidsinski A**, **Beaugerie L**, **Colombel JF**. 2004. High prevalence of adherent-invasive Escherichia coli associated with ileal mucosa in Crohn's disease. Gastroenterology.

137. **Kasai C**, **Sugimoto K**, **Moritani I**, **Tanaka J**, **Oya Y**, **Inoue H**, **Tameda M**, **Shiraki K**, **Ito M**, **Takei Y**, **Takase K**. 2016. Comparison of human gut microbiota in control subjects and patients with colorectal carcinoma in adenoma: Terminal restriction fragment length polymorphism and next-generation sequencing analyses. Oncol Rep.

138. **Cekanaviciute E**, **Yoo BB**, **Runia TF**, **Debelius JW**, **Singh S**, **Nelson CA**, **Kanner R**, **Bencosme Y**, **Lee YK**, **Hauser SL**, **Crabtree-Hartman E**, **Sand IK**, **Gacias M**, **Zhu Y**, **Casaccia P**, **Cree BAC**, **Knight R**, **Mazmanian SK**, **Baranzini SE**. 2017. Gut bacteria from multiple sclerosis patients modulate human T cells and exacerbate symptoms in mouse models. Proc Natl Acad Sci U S A.

139. **Martínez I**, **Muller CE**, **Walter J**. 2013. Long-Term Temporal Analysis of the Human Fecal Microbiota Revealed a Stable Core of Dominant Bacterial Species. PLoS One.

140. **Clarke SF**, **Murphy EF**, **O'Sullivan O**, **Ross RP**, **O'Toole PW**, **Shanahan F**, **Cotter PD**. 2013. Targeting the Microbiota to Address Diet-Induced Obesity: A Time Dependent

Challenge. PLoS One.

141. **Saulnier DM**, **Riehle K**, **Mistretta TA**, **Diaz MA**, **Mandal D**, **Raza S**, **Weidler EM**, **Qin X**, **Coarfa C**, **Milosavljevic A**, **Petrosino JF**, **Highlander S**, **Gibbs R**, **Lynch S V.**, **Shulman RJ**, **Versalovic J**. 2011. Gastrointestinal microbiome signatures of pediatric patients with irritable bowel syndrome. Gastroenterology.

142. **Zhang T**, **Breitbart M**, **Lee WH**, **Run JQ**, **Wei CL**, **Soh SWL**, **Hibberd ML**, **Liu ET**, **Rohwer F**, **Ruan Y**. 2006. RNA viral community in human feces: Prevalence of plant pathogenic viruses. PLoS Biol **4**:0108–0118.

143. **Reyes A**, **Haynes M**, **Hanson N**, **Angly FE**, **Heath AC**, **Rohwer F**, **Gordon JI**. 2010. Viruses in the faecal microbiota of monozygotic twins and their mothers. Nature **466**:334–338.

144. **Breitbart M**, **Hewson I**, **Felts B**, **Mahaffy JM**, **Nulton J**, **Salamon P**, **Rohwer F**. 2003. Metagenomic analyses of an uncultured viral community from human feces. J Bacteriol **185**:6220–6223.

145. **Roux S**, **Enault F**, **Hurwitz BL**, **Sullivan MB**. 2015. VirSorter: Mining viral signal from microbial genomic data. PeerJ **2015**:1–20.

146. **Krishnamurthy SR**, **Wang D**. 2016. Origins and challenges of viral dark matter. Virus Res.

147. **Minot S**, **Sinha R**, **Chen J**, **Li H**, **Keilbaugh SA**, **Wu GD**, **Lewis JD**, **Bushman FD**. 2011. The human gut virome: Inter-individual variation and dynamic response to diet. Genome Res **21**:1616–1625.

148. **Handley SA**, **Desai C**, **Zhao G**, **Droit L**, **Schroeder AC**, **Nkolola JP**, **Norman ME**, **Miller AD**, **Barouch DH**, **Virgin HW**, **Monaco CL**, **Schroeder AC**, **Nkolola JP**, **Norman ME**, **Miller AD**, **Wang D**, **Barouch DH**, **Virgin HW**. 2016. SIV Infection-Mediated Changes in Gastrointestinal Bacterial Microbiome and Virome Are Associated with Immunodeficiency and Prevented by Vaccination. Cell Host Microbe **19**:323–335.

149. **Reyes A**, **Blanton L V.**, **Cao S**, **Zhao G**, **Manary M**, **Trehan I**, **Smith MI**, **Wang D**, **Virgin HW**, **Rohwer F**, **Gordon JI**. 2015. Gut DNA viromes of Malawian twins discordant for severe acute malnutrition. Proc Natl Acad Sci U S A **112**:11941–11946.

150. **Minot S**, **Bryson A**. 2013. Rapid evolution of the human gut virome. Proc … **110**:12450–12455.

151. **Shkoporov AN**, **Ryan FJ**, **Draper LA**, **Forde A**, **Stockdale SR**, **Daly KM**, **McDonnell SA**, **Nolan JA**, **Sutton TDS**, **Dalmasso M**, **McCann A**, **Ross RP**, **Hill C**. 2018. Reproducible protocols for metagenomic analysis of human faecal phageomes. Microbiome **6**:68.

152. **Browne HP**, **Forster SC**, **Anonye BO**, **Kumar N**, **Neville BA**, **Stares MD**, **Goulding D**, **Lawley TD**. 2016. Culturing of "unculturable" human microbiota reveals novel taxa and extensive sporulation. Nature.

153. **Breitbart M**, **Haynes M**, **Kelley S**, **Angly F**, **Edwards RA**, **Felts B**, **Mahaffy JM**, **Mueller J**, **Nulton J**, **Rayhawk S**, **Rodriguez-Brito B**, **Salamon P**, **Rohwer F**. 2008. Viral diversity and dynamics in an infant gut. Res Microbiol **159**:367–373.

154. **Lim ES**, **Wang D**, **Holtz LR**. 2016. The Bacterial Microbiome and Virome Milestones of Infant Development. Trends Microbiol **24**:801–810.

155. **Abeles SR**, **Ly M**, **Santiago-Rodriguez TM**, **Pride DT**. 2015. Effects of long term antibiotic therapy on human oral and fecal viromes. PLoS One **10**:1–18.

156. **Norman JM**, **Handley SA**, **Baldridge MT**, **Droit L**, **Liu CY**, **Keller BC**, **Kambal A**, **Monaco CL**, **Zhao G**, **Fleshner P**, **Stappenbeck TS**, **McGovern DPB**, **Keshavarzian A**, **Mutlu EA**, **Sauk J**, **Gevers D**, **Xavier RJ**, **Wang D**, **Parkes M**, **Virgin HW**, **Norman JM**, **Handley SA**, **Baldridge MT**, **Droit L**, **Liu CY**, **Keller BC**. 2015. Disease-Specific Alterations in the Enteric Virome in Inflammatory Bowel Disease. Cell **160**:447–460.

157. **Nakatsu G**, **Zhou H**, **Ka W**, **Wu K**, **Wong SH**, **Coker OO**, **Dai Z**, **Li X**, **Szeto C**, **Sugimura N**, **Lam TY**, **Yu AC**, **Wang X**, **Chen Z**, **Wong MC**, **Ng SC**, **Tak M**, **Chan V**, **Kay P**, **Chan S**, **Ka F**, **Chan L**, **Sung JJ**, **Yu J**. 2018. Alterations in Enteric Virome Are Associated With Colorectal Cancer and Survival Outcomes. Gastroenterology **155**:529-541.e5.

158. **Kim M-S**, **Park E-J**, **Roh SW**, **Bae J-W**. 2011. Diversity and Abundance of Single-Stranded DNA Viruses in Human Feces. Appl Environ Microbiol **77**:8062–8070.

159. **Rodriguez-Brito B**, **Li L**, **Wegley L**, **Furlan M**, **Angly F**, **Breitbart M**, **Buchanan J**, **Desnues C**, **Dinsdale E**, **Edwards R**, **Felts B**, **Haynes M**, **Liu H**, **Lipson D**, **Mahaffy J**, **Martin-Cuadrado AB**, **Mira A**, **Nulton J**, **Pašić L**, **Rayhawk S**, **Rodriguez-Mueller J**, **Rodriguez-Valera F**, **Salamon P**, **Srinagesh S**, **Thingstad TF**, **Tran T**, **Thurber RV**, **Willner D**, **Youle M**, **Rohwer F**. 2010. Viral and microbial community dynamics in four aquatic environments. ISME J **4**:739–751.

160. **Thingstad T**. 1998. A theorethical approach to structuring mechanisms in the pelagial food web. Hydrobiol **3631359**:59–72.

161. **Bohannan BJM**, **Lenski RE**. 1997. Effect of resource enrichment on a chemostat community of bacteria and bacteriophage. Ecology **78**:2303–2315.

162. **Winter C**, **Bouvier T**, **Weinbauer MG**, **Thingstad TF**. 2010. Trade-Offs between Competition and Defense Specialists among Unicellular Planktonic Organisms: the "Killing the Winner" Hypothesis Revisited. Microbiol Mol Biol Rev **74**:42–57.

163. **Duerkop B a**, **Clements C V**, **Rollins D**, **Rodrigues JLM**, **Hooper L V**. 2012. A composite bacteriophage alters colonization by an intestinal commensal bacterium. Proc Natl Acad Sci U S A **109**:17621–6.

164. **Cohen LW**. 1969. Delayed lysis with salmonella bacteriophage p22: induction of lysis by addition of cysteine or histidine to the growth medium. J Virol.

165. **Lunde M**, **Aastveit AH**, **Blatny JM**, **Nes IF**. 2005. Effects of diverse environmental

conditions on φLC3 prophage stability in Lactococcus lactis. Appl Environ Microbiol.

166. **Smit E**, **Wolters AC**, **Lee H**, **Trevors JT**, **Van Elsas JD**. 1996. Interactions between a genetically marked Pseudomonas fluorescens strain and bacteriophage ΦR2f in Soil: Effects of nutrients, alginate encapsulation, and the wheat rhizosphere. Microb Ecol.

167. **Williamson SJ**, **Paul JH**. 2004. Nutrient stimulation of lytic phage production in bacterial populations of the Gulf of Mexico. Aquat Microb Ecol.

168. **Lymer D**, **Lindström ES**. 2010. Changing phosphorus concentration and subsequent prophage induction alter composition of a freshwater viral assemblage. Freshw Biol.

169. **Wang J**, **Gao Y**, **Zhao F**. 2016. Phage-bacteria interaction network in human oral microbiome. Environ Microbiol **18**:2143–2158.

170. **Amarillas L**, **Chaidez C**, **González-Robles A**, **León-Félix J**. 2016. Complete genome sequence of new bacteriophage phiE142, which causes simultaneously lysis of multidrug-resistant Escherichia coli O157:H7 and Salmonella enterica. Stand Genomic Sci **11**:89.

171. **Niu YD**, **McAllister TA**, **Nash JHE**, **Kropinski AM**, **Stanford K**. 2014. Four Escherichia coli O157:H7 phages: A new bacteriophage genus and taxonomic classification of T1-like phages. PLoS One.

172. **Koskella B**, **Meaden S**. 2013. Understanding bacteriophage specificity in natural microbial communities. Viruses **5**:806–823.

173. **De Smet J**, **Zimmermann M**, **Kogadeeva M**, **Ceyssens P-J**, **Vermaelen W**, **Blasdel B**, **Bin Jang H**, **Sauer U**, **Lavigne R**. 2016. High coverage metabolomics analysis reveals phage-specific alterations to Pseudomonas aeruginosa physiology during infection. Isme J 1–13.

174. **Kimmitt PT**, **Harwood CR**, **Barer MR**. 2000. Toxin gene expression by shiga toxin-producing Escherichia coli: the role of antibiotics and the bacterial SOS response. Emerg Infect Dis **6**:458–65.

175. **Muniesa M**, **Recktenwald J**, **Bielaszewska M**, **Karch H**, **Schmidt H**. 2000. Characterization of a Shiga toxin 2e-converting bacteriophage from an Escherichia coli strain of human origin. Infect Immun **68**:4850–4855.

176. **Hannigan GD**, **Meisel JS**, **Tyldsley AS**, **Zheng Q**, **Hodkinson BP**, **Sanmiguel AJ**, **Minot S**, **Bushman FD**, **Grice EA**. 2015. The Human Skin Double-Stranded DNA Virome: Topographical and Temporal Diversity, Genetic Enrichment, and Dynamic Associations with the Host Microbiome. MBio **6**:1–13.

177. **Modi SR**, **Lee HH**, **Spina CS**, **Collins JJ**. 2013. Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome. Nature.

178. **Moreno-Gallego JL**, **Chou SP**, **Di Rienzi SC**, **Goodrich JK**, **Spector TD**, **Bell JT**, **Youngblut ND**, **Hewson I**, **Reyes A**, **Ley RE**. 2019. Virome Diversity Correlates with Intestinal Microbiome Diversity in Adult Monozygotic Twins. Cell Host Microbe **25**:261-

272.e5.

179. **Reyes A**, **Wu M**, **McNulty NP**, **Rohwer FL**, **Gordon JI**. 2013. Gnotobiotic mouse model of phage-bacterial host dynamics in the human gut. Proc Natl Acad Sci U S A **110**:20236–41.

180. **Nguyen S**, **Baker K**, **Padman BS**, **Patwa R**, **Dunstan RA**, **Weston TA**, **Schlosser K**, **Bailey B**, **Lithgow T**, **Lazarou M**, **Luque A**, **Rohwer F**, **Blumberg RS**, **Barr JJ**. 2017. Bacteriophage Transcytosis Provides a Mechanism To Cross Epithelial Cell Layers. MBio **8**:e01874-17.

181. **Barr JJ**, **Auro R**, **Furlan M**, **Whiteson KL**, **Erb ML**, **Pogliano J**, **Stotland A**, **Wolkowicz R**, **Cutting AS**, **Doran KS**, **Salamon P**, **Youle M**, **Rohwer F**. 2013. Bacteriophage adhering to mucus provide a non-host-derived immunity. Proc Natl Acad Sci U S A **110**:10771–6.

182. **Zoetendal EG**, **Raes J**, **Van Den Bogert B**, **Arumugam M**, **Booijink CC**, **Troost FJ**, **Bork P**, **Wels M**, **De Vos WM**, **Kleerebezem M**. 2012. The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates. ISME J.

183. **Knowles B**, **Silveira CB**, **Bailey BA**, **Barott K**, **Cantu VA**, **Cobian-Guëmes AG**, **Coutinho FH**, **Dinsdale EA**, **Felts B**, **Furby KA**, **George EE**, **Green KT**, **Gregoracci GB**, **Haas AF**, **Haggerty JM**, **Hester ER**, **Hisakawa N**, **Kelly LW**, **Lim YW**, **Little M**, **Luque A**, **McDole-Somera T**, **McNair K**, **De Oliveira LS**, **Quistad SD**, **Robinett NL**, **Sala E**, **Salamon P**, **Sanchez SE**, **Sandin S**, **Silva GGZ**, **Smith J**, **Sullivan C**, **Thompson C**, **Vermeij MJA**, **Youle M**, **Young C**, **Zgliczynski B**, **Brainard R**, **Edwards RA**, **Nulton J**, **Thompson F**, **Rohwer F**. 2016. Lytic to temperate switching of viral communities. Nature.

184. **Silveira CB**, **Rohwer FL**. 2016. Piggyback-The-Winner in host-Associated microbial Communities. npj Biofilms Microbiomes.

185. **Górski A**, **Dąbrowska K**, **Międzybrodzki R**, **Weber-Dąbrowska B**, **Łusiak-Szelachowska M**, **Jończyk-Matysiak E**, **Borysowski J**. 2017. Phages and immunomodulation. Future Microbiol **12**:905–914.

186. **Duerkop BA**, **Hooper L V.** 2013. Resident viruses and their interactions with the immune system. Nat Immunol.

187. **Duerkop BA**, **Kleiner M**, **Paez-Espino D**, **Zhu W**, **Bushnell B**, **Hassell B**, **Winter SE**, **Kyrpides NC**, **Hooper L V.** 2018. Murine colitis reveals a disease-associated bacteriophage community. Nat Microbiol.

188. **Gogokhia L**, **Buhrke K**, **Bell R**, **Hoffman B**, **Brown DG**, **Hanke-Gogokhia C**, **Ajami NJ**, **Wong MC**, **Ghazaryan A**, **Valentine JF**, **Porter N**, **Martens E**, **O'Connell R**, **Jacob V**, **Scherl E**, **Crawford C**, **Stephens WZ**, **Casjens SR**, **Longman RS**, **Round JL**. 2019. Expansion of Bacteriophages Is Linked to Aggravated Intestinal Inflammation and Colitis. Cell Host Microbe **25**:285-299.e8.

189. **Broecker F**, **Klumpp J**, **Schuppler M**, **Russo G**, **Biedermann L**, **Hombach M**, **Rogler G**, **Moelling K**. 2016. Long-term changes of bacterial and viral compositions in the intestine of a recovered Clostridium difficile patient after fecal microbiota transplantation . Mol Case Stud **2**:a000448.

190. **Broecker F**, **Russo G**, **Klumpp J**, **Moelling K**. 2016. Stable core virome despite variable microbiome after fecal transfer. Gut Microbes **8**:1–7.

191. **Pride DT**, **Salzman J**, **Haynes M**, **Rohwer F**, **Davis-Long C**, **White 3rd RA**, **Loomer P**, **Armitage GC**, **Relman DA**. 2012. Evidence of a robust resident bacteriophage population revealed through analysis of the human salivary virome. Isme J **6**:915–926.

192. **Manrique P**, **Bolduc B**, **Walk ST**, **Van Oost J Der**, **De Vos WM**, **Young MJ**. 2016. Healthy human gut phageome. Proc Natl Acad Sci U S A **113**:10400–10405.

193. **Liang YY**, **Zhang W**, **Tong YG**, **Chen SP**. 2016. crAssphage is not associated with diarrhoea and has high genetic diversity. Epidemiol Infect **144**:3549–3553.

194. **Fernandes MA**, **Verstraete SG**, **Phan TG**, **Deng X**, **Stekol E**, **LaMere B**, **Lynch S V.**, **Heyman MB**, **Delwart E**. 2019. Enteric Virome and Bacterial Microbiota in Children With Ulcerative Colitis and Crohn Disease. J Pediatr Gastroenterol Nutr **68**:30–36.

195. **McCann A**, **Ryan FJ**, **Stockdale SR**, **Dalmasso M**, **Blake T**, **Anthony Ryan C**, **Stanton C**, **Mills S**, **Ross PR**, **Hill C**. 2018. Viromes of one year old infants reveal the impact of birth mode on microbiome diversity. PeerJ **2018**:1–13.

196. **Roux S**, **Krupovic M**, **Poulet A**, **Debroas D**, **Enault F**. 2012. Evolution and diversity of the microviridae viral family through a collection of 81 new complete genomes assembled from virome reads. PLoS One.

197. **Krupovic M**, **Forterre P**. 2011. Microviridae goes temperate: Microvirus-related proviruses reside in the genomes of bacteroidetes. PLoS One.

198. **Kim KH**, **Chang HW**, **Nam Y Do**, **Roh SW**, **Kim MS**, **Sung Y**, **Jeon CO**, **Oh HM**, **Bae JW**. 2008. Amplification of uncultured single-stranded DNA viruses from rice paddy soil. Appl Environ Microbiol.

199. **Kim KH**, **Bae JW**. 2011. Amplification methods bias metagenomic libraries of uncultured single-stranded and double-stranded DNA viruses. Appl Environ Microbiol.

200. **Victoria JG**, **Kapoor A**, **Li L**, **Blinkova O**, **Slikas B**, **Wang C**, **Naeem A**, **Zaidi S**, **Delwart E**. 2009. Metagenomic Analyses of Viruses in Stool Samples from Children with Acute Flaccid Paralysis. J Virol **83**:4642–4651.

201. **Linsuwanon P**, **Poovorawan Y**, **Li L**, **Deng X**, **Vongpunsawad S**, **Delwart E**. 2015. The fecal virome of children with hand, foot, and mouth disease that tested PCR negative for pathogenic enteroviruses. PLoS One **10**:1–20.

202. **Wang C**, **Zhou S**, **Xue W**, **Shen L**, **Huang W**, **Zhang Y**, **Li X**, **Wang J**, **Zhang H**, **Ma X**. 2018. Comprehensive virome analysis reveals the complexity and diversity of the viral

spectrum in pediatric patients diagnosed with severe and mild hand-foot-and-mouth disease. Virology **518**:116–125.

203. **Finkbeiner SR**, **Allred AF**, **Tarr PI**, **Klein EJ**, **Kirkwood CD**, **Wang D**. 2008. Metagenomic analysis of human diarrhea: Viral detection and discovery. PLoS Pathog **4**.

204. **Nakamura S**, **Yang CS**, **Sakon N**, **Ueda M**, **Tougan T**, **Yamashita A**, **Goto N**, **Takahashi K**, **Yasunaga T**, **Ikuta K**, **Mizutani T**, **Okamoto Y**, **Tagami M**, **Morita R**, **Maeda N**, **Kawai J**, **Hayashizaki Y**, **Nagai Y**, **Horii T**, **Iida T**, **Nakaya T**. 2009. Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. PLoS One **4**:1–8.

205. **Phan TG**, **Vo NP**, **Bonkoungou IJO**, **Kapoor A**, **Barro N**, **O'Ryan M**, **Kapusinszky B**, **Wang C**, **Delwart E**. 2012. Acute Diarrhea in West African Children: Diverse Enteric Viruses and a Novel Parvovirus Genus. J Virol **86**:11024–11030.

206. **Smits SL**, **Schapendonk CME**, **van Beek J**, **Vennema H**, **Schürch AC**, **Schipper D**, **Bodewes R**, **Haagmans BL**, **Osterhaus ADMEME**, **Koopmans MP**. 2014. New Viruses in Idiopathic Human Diarrhea Cases, the Netherlands. Emerg Infect Dis **20**:1218–1222.

207. **Bodewes R**, **van der Giessen J**, **Haagmans BL**, **Osterhaus ADME**, **Smits SL**. 2013. Identification of Multiple Novel Viruses, Including a Parvovirus and a Hepevirus, in Feces of Red Foxes. J Virol **87**:7758–7764.

208. **Moore NE**, **Wang J**, **Hewitt J**, **Croucher D**, **Williamson DA**, **Paine S**, **Yen S**, **Greening GE**, **Hall RJ**. 2015. Metagenomic analysis of viruses in feces from unsolved outbreaks of gastroenteritis in humans. J Clin Microbiol **53**:15–21.

209. **Van Leeuwen M**, **Williams MMW**, **Koraka P**, **Simon JH**, **Smits SL**, **Osterhaus ADME**. 2010. Human picobirnaviruses identified by molecular screening of diarrhea samples. J Clin Microbiol.

210. **Holtz LR**, **Cao S**, **Zhao G**, **Bauer IK**, **Denno DM**, **Klein EJ**, **Antonio M**, **Stine OC**, **Snelling TL**, **Kirkwood CD**, **Wang D**. 2014. Geographic variation in the eukaryotic virome of human diarrhea. Virology **468**:556–564.

211. **Phan TG**, **Kapusinszky B**, **Wang C**, **Rose RK**, **Lipton HL**, **Delwart EL**. 2011. The Fecal Viral Flora of Wild Rodents. PLoS Pathog **7**:e1002218.

212. **Day JM**, **Ballard LL**, **Duke M V**, **Scheffler BE**, **Zsak L**. 2010. Metagenomic analysis of the turkey gut RNA virus community. Virol J **7**:313.

213. **Yinda CK**, **Zell R**, **Deboutte W**, **Zeller M**, **Conceição-Neto N**, **Heylen E**, **Maes P**, **Knowles NJ**, **Ghogomu SM**, **Van Ranst M**, **Matthijnssens J**. 2017. Highly diverse population of Picornaviridae and other members of the Picornavirales, in Cameroonian fruit bats. BMC Genomics **18**:249.

214. **Yinda CK**, **Ghogomu SM**, **Conceição-Neto N**, **Beller L**, **Deboutte W**, **Vanhulle E**, **Maes P**, **Van Ranst M**, **Matthijnssens J**. 2018. Cameroonian fruit bats harbor divergent viruses, including rotavirus H, bastroviruses, and picobirnaviruses using an alternative genetic code.

Virus Evol.

215. **Krishnamurthy SR**, **Wang D**. 2018. Extensive conservation of prokaryotic ribosomal binding sites in known and novel picobirnaviruses. Virology.

216. **Bernardin F**, **Operskalski E**, **Busch M**, **Delwart E**. 2010. Transfusion transmission of highly prevalent commensal human viruses. Transfusion.

217. **Smits SL**, **Raj VS**, **Oduber MD**, **Schapendonk CME**, **Bodewes R**, **Provacia L**, **Stittelaar KJ**, **Osterhaus ADME**, **Haagmans BL**. 2013. Metagenomic Analysis of the Ferret Fecal Viral Flora. PLoS One **8**.

218. **Karlsson OE**, **Larsson J**, **Hayer J**, **Berg M**, **Jacobson M**. 2016. The intestinal eukaryotic virome in healthy and diarrhoeic neonatal piglets. PLoS One **11**:1–16.

219. **Hause BM**, **Padmanabhan A**, **Pedersen K**, **Gidlewski T**. 2016. Feral swine virome is dominated by single-stranded DNA viruses and contains a novel Orthopneumovirus which circulates both in feral and domestic swine. J Gen Virol **97**:2090–2095.

220. **Kluge M**, **Campos FS**, **Tavares M**, **de Amorim DB**, **Valdez FP**, **Giongo A**, **Roehe PM**, **Franco AC**. 2016. Metagenomic Survey of Viral Diversity Obtained from Feces of Subantarctic and South American Fur Seals. PLoS One **11**:e0151921.

221. **Freer G**, **Maggi F**, **Pifferi M**, **Di Cicco ME**, **Peroni DG**, **Pistello M**. 2018. The virome and its major component, Anellovirus, a convoluted system molding human immune defenses and possibly affecting the development of asthma and respiratory diseases in childhood. Front Microbiol **9**:1–7.

222. **Young JC**, **Chehoud C**, **Bittinger K**, **Bailey A**, **Diamond JM**, **Cantu E**, **Haas AR**, **Abbas A**, **Frye L**, **Christie JD**, **Bushman FD**, **Collman RG**. 2015. Viral metagenomics reveal blooms of anelloviruses in the respiratory tract of lung transplant recipients. Am J Transplant **15**:200–209.

223. **Ngoi CN**, **Siqueira J**, **Li L**, **Deng X**, **Mugo P**, **Graham SM**, **Price MA**, **Sanders EJ**, **Delwart E**. 2016. The plasma virome of febrile adult kenyans shows frequent parvovirus B19 infections and a novel arbovirus (Kadipiro virus). J Gen Virol **97**:3359–3367.

224. **Dinakaran V**, **Rathinavel A**, **Pushpanathan M**, **Sivakumar R**, **Gunasekaran P**, **Rajendhran J**. 2014. Elevated levels of circulating DNA in cardiovascular disease patients: metagenomic profiling of microbiome in the circulation. PLoS One **9**:e105221.

225. **Gootenberg DB**, **Paer JM**, **Luevano J-M**, **Kwon DS**. 2016. HIV-associated changes in the enteric microbial community. Curr Opin Infect Dis **0**:1.

226. **Lima DA**, **Cibulski SP**, **Tochetto C**, **Varela APM**, **Finkler F**, **Teixeira TF**, **Loiko MR**, **Cerva C**, **Junqueira DM**, **Mayer FQ**, **Roehe PM**. 2019. The intestinal virome of malabsorption syndrome-affected and unaffected broilers through shotgun metagenomics. Virus Res **261**:9–20.

227. **Wang J**, **Zheng J**, **Shi W**, **Du N**, **Xu X**, **Zhang Y**, **Ji P**, **Zhang F**, **Jia Z**, **Wang Y**, **Zheng**

Z, **Zhang H**, **Zhao F**. 2018. Dysbiosis of maternal and neonatal microbiota associated with gestational diabetes mellitus. Gut.

228. **Barton ES**, **White DW**, **Cathelyn JS**, **Brett-McClellan KA**, **Engle M**, **Diamond MS**, **Miller VL**, **Virgin IV HW**. 2007. Herpesvirus latency confers symbiotic protection from bacterial infection. Nature.

229. **Chatterjee A**, **Duerkop BA**. 2018. Beyond bacteria: Bacteriophage-eukaryotic host interactions reveal emerging paradigms of health and disease. Front Microbiol **9**:1–8.

230. **Kim KW**, **Horton JL**, **Pang CNI**, **Jain K**, **Leung P**, **Isaacs SR**, **Bull RA**, **Luciani F**, **Wilkins MR**, **Catteau J**, **Lipkin WI**, **Rawlinson WD**, **Briese T**, **Craig ME**. 2019. Higher abundance of enterovirus A species in the gut of children with islet autoimmunity. Sci Rep **9**:1–8.

231. **Zhao G**, **Vatanen T**, **Droit L**, **Park A**, **Kostic AD**, **Poon TW**, **Vlamakis H**, **Siljander H**, **Härkönen T**, **Hämäläinen A-M**, **Peet A**, **Tillmann V**, **Ilonen J**, **Wang D**, **Knip M**, **Xavier RJ**, **Virgin HW**. 2017. Intestinal virome changes precede autoimmunity in type I diabetes-susceptible children. Proc Natl Acad Sci 201706359.

232. **McLaughlin-Drubin ME**, **Munger K**. 2008. Viruses associated with human cancer. Biochim Biophys Acta - Mol Basis Dis.

233. **Degruttola AK**, **Low D**, **Mizoguchi A**, **Mizoguchi E**. 2016. Current understanding of dysbiosis in disease in human and animal models. Inflamm Bowel Dis **22**:1137–1150.

234. **De Cárcer DA**, **Hernáez B**, **Rastrojo A**, **Alcamí A**. 2017. Infection with diverse immune-modulating poxviruses elicits different compositional shifts in the mouse gut microbiome. PLoS One **12**:1–9.

235. **Alcami A**. 2003. Viral mimicry of cytokines, chemokines and their receptors. Nat Rev Immunol.

236. **Finlay BB**, **McFadden G**. 2006. Anti-immunology: Evasion of the host immune system by bacterial and viral pathogens. Cell.

237. **Kernbauer E**, **Ding Y**, **Cadwell K**. 2014. An enteric virus can replace the beneficial function of commensal bacteria. Nature **516**:94–98.

238. **Yang JY**, **Kim MS**, **Kim E**, **Cheon JH**, **Lee YS**, **Kim Y**, **Lee SH**, **Seo SU**, **Shin SH**, **Choi SS**, **Kim B**, **Chang SY**, **Ko HJ**, **Bae JW**, **Kweon MN**. 2016. Enteric Viruses Ameliorate Gut Inflammation via Toll-like Receptor 3 and Toll-like Receptor 7-Mediated Interferon-β Production. Immunity **44**:889–900.

239. **Chen S**, **Tsai C**, **Lee Y**, **Lin C**, **Huang K**. 2017. Intestinal microbiome in children with severe and complicated acute viral gastroenteritis. Nat Publ Gr 1–7.

240. **Nelson AM**, **Walk ST**, **Taube S**, **Taniuchi M**, **Houpt ER**, **Wobus CE**, **Young VB**. 2012. Disruption of the Human Gut Microbiota following Norovirus Infection. PLoS One **7**.

241. **Youmans BP**, **Ajami NJ**, **Jiang Z-D**, **Campbell F**, **Wadsworth WD**, **Petrosino JF**, **DuPont HL**, **Highlander SK**. 2015. Characterization of the human gut microbiome during travelers' diarrhea. Gut Microbes **6**:110–119.

242. **Chen S-Y**, **Tsai C-N**, **Lee Y-S**, **Lin C-Y**, **Huang K-Y**, **Chao H-C**, **Lai M-W**, **Chiu C-H**. 2017. Intestinal microbiome in children with severe and complicated acute viral gastroenteritis. Sci Rep **7**:46130.

243. **Aiemjoy K**, **Altan E**, **Aragie S**, **Fry DM**, **Phan TG**, **Deng X**, **Chanyalew M**, **Tadesse Z**, **Callahan EK**, **Delwart E**, **Keenan JD**. 2019. Viral species richness and composition in young children with loose or watery stool in Ethiopia. BMC Infect Dis **19**:53.

244. **Vandeputte D**, **Falony G**, **Vieira-Silva S**, **Tito RY**, **Joossens M**, **Raes J**. 2016. Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates. Gut **65**:57–62.

245. **Hendaus MA**, **Jomha FA**, **Alhammadi AH**. 2015. Virus-induced secondary bacterial infection: A concise review. Ther Clin Risk Manag **11**:1265–1271.

246. **Mathew S**, **Smatti MK**, **Al Ansari K**, **Nasrallah GK**, **Al Thani AA**, **Yassine HM**. 2019. Mixed Viral-Bacterial Infections and Their Effects on Gut Microbiota and Clinical Illnesses in Children. Sci Rep **9**:1–12.

247. **Stecher B**, **Robbiani R**, **Walker AW**, **Westendorf AM**, **Barthel M**, **Kremer M**, **Chaffron S**, **Macpherson AJ**, **Buer J**, **Parkhill J**, **Dougan G**, **Von Mering C**, **Hardt WD**. 2007. Salmonella enterica serovar typhimurium exploits inflammation to compete with the intestinal microbiota. PLoS Biol.

248. **Arumugam M**, **Raes J**, **Pelletier E**, **Le Paslier D**, **Yamada T**, **Mende DR**, **Fernandes GR**, **Tap J**, **Bruls T**, **Batto JM**, **Bertalan M**, **Borruel N**, **Casellas F**, **Fernandez L**, **Gautier L**, **Hansen T**, **Hattori M**, **Hayashi T**, **Kleerebezem M**, **Kurokawa K**, **Leclerc M**, **Levenez F**, **Manichanh C**, **Nielsen HB**, **Nielsen T**, **Pons N**, **Poulain J**, **Qin J**, **Sicheritz-Ponten T**, **Tims S**, **Torrents D**, **Ugarte E**, **Zoetendal EG**, **Wang J**, **Guarner F**, **Pedersen O**, **De Vos WM**, **Brunak S**, **Doré J**, **Weissenbach J**, **Ehrlich SD**, **Bork P**. 2011. Enterotypes of the human gut microbiome. Nature.

249. **Boutard M**, **Cerisy T**, **Nogue PY**, **Alberti A**, **Weissenbach J**, **Salanoubat M**, **Tolonen AC**. 2014. Functional Diversity of Carbohydrate-Active Enzymes Enabling a Bacterium to Ferment Plant Biomass. PLoS Genet.

250. **Peterson DA**, **Frank DN**, **Pace NR**, **Gordon JI**. 2008. Metagenomic Approaches for Defining the Pathogenesis of Inflammatory Bowel Diseases. Cell Host Microbe.

251. **Sokol H**, **Pigneur B**, **Watterlot L**, **Lakhdari O**, **Bermúdez-Humarán LG**, **Gratadoux JJ**, **Blugeon S**, **Bridonneau C**, **Furet JP**, **Corthier G**, **Grangette C**, **Vasquez N**, **Pochart P**, **Trugnan G**, **Thomas G**, **Blottière HM**, **Doré J**, **Marteau P**, **Seksik P**, **Langella P**. 2008. Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. Proc Natl Acad Sci U S A.

252. **Verdam FJ**, **Fuentes S**, **De Jonge C**, **Zoetendal EG**, **Erbil R**, **Greve JW**, **Buurman WA**, **De Vos WM**, **Rensen SS**. 2013. Human intestinal microbiota composition is associated with local and systemic inflammation in obesity. Obesity.

253. **Finegold S**, **Summanen P**, **Hunt Gerardo S**, **Baron E**. 1992. Clinical importance of Bilophila wadsworthia. Eur J Clin Microbiol Infect Dis.

254. **Abdulamir AS**, **Hafidh RR**, **Bakar FA**. 2011. The association of Streptococcus bovis/gallolyticus with colorectal tumors: The nature and the underlying mechanisms of its etiological role. J Exp Clin Cancer Res.

255. **Qiao Y**, **Sun J**, **Xie Z**, **Shi Y**, **Le G**. 2014. Propensity to high-fat diet-induced obesity in mice is associated with the indigenous opportunistic bacteria on the interior of Peyer's patches. J Clin Biochem Nutr.

256. **Segata N**, **Haake SK**, **Mannon P**, **Lemon KP**, **Waldron L**, **Gevers D**, **Huttenhower C**, **Izard J**. 2012. Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. Genome Biol.

257. **Wei X**, **Yan X**, **Zou D**, **Yang Z**, **Wang X**, **Liu W**, **Wang S**, **Li X**, **Han J**, **Huang L**, **Yuan J**. 2013. Abnormal fecal microbiota community and functions in patients with hepatitis B liver cirrhosis as revealed by a metagenomic approach. BMC Gastroenterol.

258. **Morgan XC**, **Tickle TL**, **Sokol H**, **Gevers D**, **Devaney KL**, **Ward D V**, **Reyes JA**, **Shah SA**, **LeLeiko N**, **Snapper SB**, **Bousvaros A**, **Korzenik J**, **Sands BE**, **Xavier RJ**, **Huttenhower C**. 2012. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol **13**:R79.

259. **Rooks MG**, **Veiga P**, **Wardwell-Scott LH**, **Tickle T**, **Segata N**, **Michaud M**, **Gallini CA**, **Beal C**, **Van Hylckama-Vlieg JET**, **Ballal SA**, **Morgan XC**, **Glickman JN**, **Gevers D**, **Huttenhower C**, **Garrett WS**. 2014. Gut microbiome composition and function in experimental colitis during active disease and treatment-induced remission. ISME J.

260. **d'Hennezel E**, **Abubucker S**, **Murphy LO**, **Cullen TW**. 2017. Total Lipopolysaccharide from the Human Gut Microbiome Silences Toll-Like Receptor Signaling. mSystems.

261. **Loo YM**, **Gale M**. 2011. Immune Signaling by RIG-I-like Receptors. Immunity.

262. **Formosa-Dague C**, **Castelain M**, **Martin-Yken H**, **Dunker K**, **Dague E**, **Sletmoen M**. 2018. The Role of Glycans in Bacterial Adhesion to Mucosal Surfaces: How Can Single-Molecule Techniques Advance Our Understanding? Microorganisms.

263. **Gupta P**, **Song B**, **Neto C**, **Camesano TA**. 2016. Atomic force microscopy-guided fractionation reveals the influence of cranberry phytochemicals on adhesion of Escherichia coli. Food Funct.

264. **Ivey KL**, **Chan AT**, **Izard J**, **Cassidy A**, **Rogers GB**, **Rimma EB**. 2019. Role of dietary flavonoid compounds in driving patterns of microbial community assembly. MBio.

265. **Abeles SR**, **Robles-Sikisaka R**, **Ly M**, **Lum AG**, **Salzman J**, **Boehm TK**, **Pride DT**. 2014.

Human oral viruses are personal, persistent and gender-consistent. ISME J **8**:1753–1767.

266. **Hillman ET**, **Lu H**, **Yao T**, **Nakatsu CH**. 2017. Microbial ecology along the gastrointestinal tract. Microbes Environ.

CHAPTER 2

ASSOCIATIONS BETWEEN VIROME AND BACTERIOME PROFILES AND ACUTE GASTROENTERITIS AMONG MICHIGAN PATIENTS AND HEALTHY FAMILY MEMBERS

ABSTRACT

Gastroenteritis contributes to a significant disease burden worldwide, and, while affecting all age groups, it predominately impacts children. Gastroenteritis is primarily an acute infection but can also be an inciting event to chronic diseases like inflammatory bowel disease (IBD) or irritable bowel syndrome (IBS). Gastroenteritis microbiome studies have traditionally focused on alterations in resident bacterial populations, often ignoring the distribution of viruses or the virome. Prior studies have identified Proteobacteria, specifically, the genera of *Escherichia* as being associated with gastroenteritis illness. As a result, this study was designed using metagenomics to investigate the microbiome among 79 patients with acute bacterial gastroenteritis for comparison to 125 healthy family members (controls). In total, over 1,000,000,000 reads (621,384,080 paired reads) were sequenced and evaluated. Our findings further confirm the presence of Proteobacteria-dominant microbial communities in gastroenteritis patients. We also identify disease-specific changes in the microbiome specific to infection status, which include alterations in viruses and bacteria. Two case-dominated clusters with similar microbial profiles were identified. One of the clusters (Cluster 2) was significantly associated with more severe disease and had lower diversity and richness as well as a more dysbiotic microbial profile relative to communities representing the other clusters. Cluster 2 had 82 differentially abundant genera compared to other clusters, as was identified using Analysis of Composition of Microbiomes (ANCOM); 26 genera were above average in abundance in Cluster 2 relative to the average of all samples across the study for a given genus. Further analysis of these 26 genera using logistic regression identified four genera (*Acinetobacteria, Salmonella, Orthopoxvirus, Serratia*) to be strong features of Cluster 2 status. Identification of the microbes presented here builds on the understanding of enteric infections and could help identify novel avenues to therapy.

INTRODUCTION

Acute gastroenteritis maintains a significant health burden globally; the estimated incidence of acute gastroenteritis in the United States ranges from 179 million (1) to 375 million cases (2), with many cases going unreported. Children are disproportionally affected; there are 1.5 million office visits, 200,000 hospitalizations, and 300 deaths directly attributable to acute gastroenteritis in the United States (3). Worldwide there are 2.3 billion cases of acute gastroenteritis and 1.3 million deaths annually (4). Developing countries are impacted more severely by acute gastroenteritis, and children in these countries suffer the most significant disease burden. Diarrheal illness contributes to one in eight deaths in children younger than five years (5). Causative organisms of gastroenteritis vary and include viruses (*Rotavirus*, *Norovirus*, *Adenovirus*), protozoan (*Cryptosporidium*), and bacteria (*Campylobacter jejuni*, *Escherichia coli*, *Salmonella* spp., *Shigella* spp.) (6). Bacterial agents account for greater than 50% of the disease burden globally (6). The identification of a pathogen occurs in about 50% of symptomatic cases (7). This number may be lower; however, as a recent study of 196 hospitalized cases found that only 10% were culture-positive for causative agents of bacterial illness (8). Despite this limitation, culturing remains standard for laboratory diagnosis in gastroenteritis, and decisions on how to treat these infections are based primarily on clinical presentation and culture results (9).

Amplicon sequencing has been utilized extensively to study the changes that occur in the resident microbial population within the human gut during gastroenteritis. Previously, acute bacterial infections were found to cause an increase in Proteobacteria, specifically the population of *Escherichia* (10, 11) with decreases in Firmicutes and Bacteroidetes (11). Traditional studies have focused on describing and characterizing intestinal bacterial communities due to the ease of 16S rRNA sequencing; hence, little is known about the viral communities in the gut, particularly

during infection. Defining the viral populations inhabiting the human gut is essential for a thorough understanding of disease progression in bacterial gastroenteritis. Viruses can directly cause illness (i.e., *Rotavirus*) or assist in the disease process by carrying toxin genes (e.g., Shiga toxin-encoding bacteriophage (12, 13)). Secondary infections due to viruses are a common occurrence in respiratory tract infections (14) but remain wholly understudied in gastrointestinal illness. The work presented here aims to comprehensively define both the viral and bacterial communities in 79 patients and 125 healthy family members using metagenomics with Illumina shotgun sequencing. We hypothesized that viruses (Siphoviridae and Podoviridae) that commonly infect members of Proteobacteria will be more abundant in patients than healthy individuals and will be associated with more severe disease. This work will increase the understanding of the microbial communities in gastroenteritis, which could lead to identification of novel therapeutic targets.

MATERIALS AND METHODS

Sample selection and sequencing

Samples were collected through the Enteric Research Investigative Network (ERIN) at Michigan State University (11). In brief, the ERIN study was an active surveillance system coordinated with the Michigan Department of Health and Human Services (MDHHS) and four hospitals, as described previously (11). A subset of 204 stool samples was analyzed for this study; 79 samples were collected from patients with enteric infections caused by *C. jejuni*, *Salmonella*, *Shigella*, or Shiga toxin-producing *E. coli* (STEC), and 125 were received from healthy family members of the patients (controls). All samples were placed in Cary Blair transport media, homogenized, centrifuged, and stored in triplicate at -80 °C. DNA was extracted using the QIAmp DNA Stool Mini Kit (QIAGEN; Valencia, CA). Epidemiological data, including clinical details, exposures, and demographics were extracted for each patient using the Michigan Disease Surveillance System (MDSS), while questionnaires were used to obtain data about the healthy family members. All protocols were approved by the Institutional Review Boards at Michigan State University (MSU; IRB #10-736SM) and MDHHS (842-PHALAB) as well as the four participating hospitals.

Sequencing libraries were prepared using the Illumina TruSeq Nano DNA Library Preparation Kit on a Perkin Elmer Sciclone NGS robot following the manufacturer's recommendations. Four equimolar library pools were generated with samples added in duplicate for each sequencing run. Libraries were quality controlled with qPCR and quantified with a Qubit dsDNA HS (Thermo Fisher Scientific, Waltham, MA, USA) and Caliper LabChipGX HS DNA (Caliper Life Sciences, Hopkinton, MA, USA). The library for Run 1 was loaded in two lanes of an Illumina HiSeq 2500 Rapid Run flow cell (v1) and sequenced in a 2x150bp paired-end format

using Rapid SBS reagents. The libraries for Runs 2, 3, 4, however, were each combined into separate pools, loaded onto two lanes of an Illumina HiSeq 2500 Rapid Run flow cell (v2), and sequenced in a 2x250 bp paired-end format. Base-calling was performed with the Illumina Real-Time Analysis (v1.18.61), and the output was demultiplexed and converted to FastQ format by Illumina Bcl2Fastq (v1.8.4). A non-parametric multivariate analysis of variance (NPMANOVA) test demonstrated that there was no difference ($p = 0.159$) between sequencing formats, and that library runs were homogenous via betadisper ($p = 0.715$). Similarly, the principal component analysis (PCA) showed no differences in the clustering of samples from each run (Figure 2.1A-D), and hence, the sequences were merged into a single dataset for subsequent analyses.

Power analysis

The pwr package (15) in R (16), based on Cohen's equations (17), was used to determine the necessary sample size for all statistical tests employed in this study, including Chi-square, analysis of variance (ANOVA), correlation, and regression. For all calculations, standard statistical assumptions were made ($p = 0.05$, power = 0.8, effect size = 0.5). Power curves were generated to represent the relationship between effect size and sample size. Multiple levels of power were assessed and visualized with different curves (Figure 2.2). The target power for this study was 0.8, which resulted in a sample size of 204 with an effect size ≤ 0.18. We have adequate power to detect differences between study groups (cases and controls) because we have 204 samples included.

Sequence processing and metagenomics

Processing and annotation Sequencing adaptors and low-quality reads were removed using Trimmomatic (18). FastQC (19) was used to read FastQ files and generate a quality control report that includes poor quality reads, adaptors, and GC bias. Using the methodology based on Norman *et al.* (20) and KBase (21), reads passing quality control (per base sequence quality > 30) were

compared to a database of human RefSeq genomes (GRCh38_1118, downloaded November 2018) available at the National Center for Biotechnology Information (NCBI) using Bowtie 2 (22) and SAMtools (23) to remove reads that match the human genome. Kaiju (24) annotated quality-controlled reads to generate a microbiome profile by comparing each read to a non-redundant protein database (25) of viruses, bacteria, and fungi (nr_euk, downloaded January 2019) in NCBI. The tradeoff is that as quality-controlled reads are annotated, there is less specificity in categorizing those reads at lower taxonomical levels (i.e., species). On average, 90% of reads were annotated at the Phylum level, 62% at Genus, but only 22% of reads achieved Species-level determination (Figure 2.3). The results of the Kaiju output were merged into a table of samples with corresponding taxonomical classifications using a custom python script (26). This script is identical to the *kaiju2table* function in Kaiju, but the script was created before the functionality was available. An additional script was used that parsed Kaiju output at different taxonomical levels (Phylum, Class, Order, Family, Genus, Species) and split the output into viral and non-viral annotations. The analysis was also conducted at the levels of taxonomy listed above as done previously (27).

Assemblies provide the most accurate picture for annotation and could allow for inference of genomic features. Assemblies were performed with Metaspades (28). On average, 10% of reads in cases, and 14% of reads in controls did not map to the assemblies, which was statistically significant (Mann Whitney U test $p = 0.0004845$). The significant differences in mapping frequencies of assemblies of cases and controls are a concern for introducing bias into the dataset, and thus, assemblies were not utilized for subsequent analyses. Case status differences were preserved across taxonomical levels and visualized with a Principle Components Analysis (PCA). Cases clustered distinct from controls at the Class (Figure 2.4A), Family (Figure 2.4B), Genus

(Figure 2.4C) and, Species (Figure 2.4D) level. Statistical trends were examined between sequencing depth, coverage, and alpha diversity metrics to determine if the minimum sequencing depth cutoff was adequate. Importantly, no trends were noted between the sequencing depth, coverage, and alpha diversity ($R^2 < 0.7$, spearman $p > 0.05$). The lack of statistical associations between these factors demonstrates that the minimum sequencing cutoff was enough.

Among all 204 samples, the maximum number of reads (paired-forward) sequenced in a sample was 7,427,518 (3.7 Giga base pairs [Gbp]) and the average sequencing per sample was 3,046,000 reads (1.2 Gbp). There was no significant difference between the sequencing depth for cases versus controls (Mann Whitney U test $p = 0.1886$). Rarefy (29, 30), which involves subsampling the existing dataset, normalizes the sequencing data, but has been shown to introduce bias into a metagenomic dataset (31) and hence, was not utilized in our analysis. Instead, rarefaction (32), which measures species richness, was used, and rarefaction curves (33, 34) of genera data were generated with the *rarefy* and *speccacum* functions from the vegan package in R (16). Species accumulation curves (random sampling, Figure 2.5A) and rarefaction, Figure 2.5B) achieved plateau, suggesting that sequencing depth was sufficient for both cases and controls (35). Coverage for each sample was calculated based on Nonparielle3 (36) that uses read redundancy in the sample to calculate coverage. Nonparielle3 estimated the average coverage for all samples to be 78%. The Genus-level classification was used for analysis because the sequencing depth and taxonomical information available were optimal compared to other taxonomical levels.

Cluster analysis To account for spurious associations, microbial taxa that were not present in at least 1% of samples were removed to reduce the false-positive rate of taxa significance as recommended (27, 37). Multiplicative simple replacement using the zCompositions package (38) in R (39) was used to replace zero counts in the taxonomy table, while MixOmics (37) was used

to calculate the relative abundance as a percent of the total annotated viral and bacterial populations, based on previously published methods (40). A center-log-ratio transformation was performed, and a compositional-data-analysis approach was used (41). Hierarchical clustering was performed using Ward's linkage and Aitchison distance. (42). Statistical power was considered in the selection of the optimal number of clusters. As the number of clusters increases, the statistical power will decrease for the cluster due to a smaller sample size per cluster. For example, the calculated power with the sample size for six clusters was 0.75, which makes identifying statistical associations problematic. Only clusters that retained statistical power ($\geq$ 0.8) were considered further. Finally, the distribution of cases and controls within each cluster was examined to create a balanced study design. Four clusters were determined optimal for this dataset based on the above considerations. To determine if microbial profiles were different based on clusters, a one-way NP-MANOVA was performed with the adonis function from vegan (43). The p-values calculated for multiple-hypothesis testing were adjusted using a Bonferroni correction with the *p.adjust function*. Group heterogeneity for each cluster was assessed using the *betadisper function* from the vegan package. ANCOM (44) was used to determine the differentially abundant taxa found between clusters, while SparrC (45) correlated different taxa with one another to create a taxonomical network visualized with SpiecEasi (45). The vegan package was used to calculate the alpha diversity (Shannon index), Richness (total number), and Evenness (distribution) at the genus level based on the read count of each taxonomical assignment.

Data analysis

Demographic and epidemiological data were managed using Microsoft Excel and Access. Statistical analyses were performed in R and EpiInfo (46). Chi-square and Fisher's exact tests (counts $\leq$ 5) were used to identify associations between exposure (independent) and outcome

(dependent) variables in univariate analysis; p-value < 0.05 was considered significant. Epidemiologic and demographic data were used as exposure variables to identify associations with outcomes (e.g., case status, cluster status). Clusters defined by hierarchical clustering were used as the outcome variable. Other factors, including demographics, diet, medications and travel history as well as differentially abundant microbes, were examined to identify associations with specific microbiome profiles or clusters.

Univariate variables with strong associations ($p \leq 2.0$) with outcomes of interest were included in the multivariate logistic regression model. This stepwise model was generated using forward and backward selection and specific variables such as age, sex, race, residence, and infection type, were included in the model and evaluated for confounding. Factors were added or removed if they provided significant changes in the model ($p < 0.05$), and each factor was assessed for collinearity. The Wald test was used to assess the statistical significance of each coefficient present in the model, while the Hosmer-Lemeshow test (47) was employed to assess the goodness of fit. All scripts are available at githib.com/BrianNo.

RESULTS

Characteristics of the study population

Among the 79 patients with diarrheal infections (cases), 48.1% (n=38) were males and 51.9% (n=41) were female (Table 2.2.). The highest frequency of cases occurred in the 19-64 age group (n=33, 41.8%) followed by children in the 0 and 9 age group (n=21, 26.6%). Cases resided in multiple counties throughout Michigan, although most were from Ingham (n=16, 20.51%), Wayne (n=16, 20.51%), and Washtenaw (n=11, 14.1%). A total of 48.7% (n=38) of the cases resided in urban counties, while 51.3% (n=40) were from rural areas. Most cases reported body aches (n=73, 94.8%), yet 69.1% (n=47) also reported fever and vomiting. A subset of 29 cases (37.7%) was hospitalized, and among these cases, 15 (53.6%) were hospitalized for more than two days (Table 2.3.).

The 125 otherwise healthy family members (non-infected controls) had similar characteristics when compared to the cases. Sixty-seven (54.5%) healthy individuals were male and 45.5% (n=56) were female, and most were between 19 and 64 years of age (n=64, 42%) followed by 0 to 9 years of age (n=33, 26.8%). Non-infected controls also resided in multiple locations throughout Michigan, though most lived in Oakland (n=20; 17.2%), Wayne (n=20; 17.2%), Ingham (n=15, 12.9%), and Eaton (n=13, 11.2%) counties. Approximately 52.6% (n=61) resided in urban counties, whereas 47.4% (n=55) were from rural areas (Table 2.2.).

Cases and controls had different viral and bacterial read counts

In total, 621,384,080 (284.7 Gbp) paired forward reads were sequenced across all 204 samples, yielding 3,046,000 or 1.4 Gbp paired-forward reads per sample. Cases and controls achieved average sequencing depths of 3,032,694 reads (1.4 Gbp) and 3,054,410 reads (1.4 Gbp), respectively, with no difference between study groups (Mann Whitney U test $p = 0.1886$). The

average coverage, as determined by Nonpareil3 (36), was 78% across all samples. Although cases had lower coverage (77%) than controls (78.6%), the difference was not statistically significant (Mann Whitney U test p = 0.1936). On average, across all samples, 14.2% of reads fell below quality filtering parameters. More reads were removed from control sequences (14.7%) compared to cases (13.3%), though this difference was also not significant (Mann Whitney U test p = 0.1113).

On average, 6% of quality-controlled reads were annotated as human across all samples. The abundance of human DNA differed by case status; cases contained 15.2% human reads compared to only 0.1% in controls, which was statistically significant (Mann Whitney U test p = 8.509e-05). Kaiju annotated 61.7% of reads to the Genus level that passed quality control (i.e., trimming and human read removal steps) across all samples. Controls achieved a higher annotation frequency (63.3%) compared to cases (59.3%), but the difference in frequencies was not significant (Mann Whitney U test p = 0.07632). On average, 61.3% of reads were annotated to bacteria across all samples at the Genus level, and 0.5% of reads were assigned to viruses. Cases had a lower proportion of reads assigned to bacteria (58.7%) compared to controls (62.9%; Mann Whitney U test p = 0.04888). Cases also had an increased proportion of viruses (0.7%) compared to controls (0.3%), which was statistically significant (Mann Whitney U test p = 2.45e-05) (Table 2.1.). Case communities had a lower Shannon index for diversity (Mann Whitney U test p = 0.006634, Figure 2.6A) and richness (Mann Whitney U test p = 3.212e-12, Figure 2.6B) when compared to non-infected communities at genus level. Evenness was not significantly different between cases and controls (Mann Whitney U test p = 0.1474, Figure 2.6C).

Microbiome composition varies between cases and controls

In total, 473 (449 bacterial, 24 viral) Families were identified. At the Genus level, there were 2,659 genera identified (2,482 bacteria and 177 viruses). Examination of the top five virus

Families (Figure 2.7A) between cases and controls shows that Myoviridae and Poxviridae are more abundant in cases comprising 26% and 9% of viral reads on average respectively. Poxviridae was statistically significantly higher in cases compared to controls (Mann Whitney U test p = 1.4e-12). Microviridae and Siphoviridae are more abundant in controls than cases comprising 18% and 41% of the control virome on average, respectively. Microviridae was statistically significantly lower in cases compared to controls (Mann Whitney U test p = 6.4e-10). Bacteria profiles were different as well (Figure 2.7B). Examination of the top 10 Bacteria families shows that Enterobacteriaceae are significantly more abundant in case with bacterial reads accounting for 34.4% of the total bacterial reads in cases on average, which was statistically significant from controls (2.7%, Mann Whitney U test p = 2e-16). Bacteroidaceae, Ruminococcaceae, Rikenellaceae, and Prevotellacea were all significantly more abundant in controls on average than cases, accounting for 45%, 10%, 9%, and 7%, respectively (Mann Whitney U test p = 0.01284, 1.8e-10, 1.5e-14, 1.152e-09). Collectively, these data provide support for the differences in microbiome profiles identified between patients with acute gastroenteritis and non-infected individuals (Figure 2.6, Figure 2.7).

Further analysis of the microbiome in the 79 cases identified differences in the virome by infection type. The abundance of *P22virus*, for example, was significantly different across the four infection types (Kruskal wallis p = 0.01906) and was significantly higher in the *Salmonella* cases. Among these *Salmonella* cases, *P22virus* comprised an average of 14% of viral reads across samples compared to all other infection types combined (Mann Whitney U test p = 0.02669). A similar difference was observed for *P2virus*, which comprised, on average, 11% of viral reads in the communities from Salmonella patients compared the non-*Salmonella* infections (Mann Whitney U test p = 0.04498). In *Shigella* infections, *P1virus* comprised 12% of viral reads and was statistically significant compared to non-*Shigella* infections (Mann Whitney U test p =

0.03276). No differences, however, were observed in bacteriophage populations among cases with *Campylobacter* infections.

In STEC-infected communities, *Nona33virus* comprised 20% of viral reads and was significantly different when compared to non-STEC infections (Mann Whitney U test p = 0.03096). Patients with Shigella (11%) and Salmonella (4%) infections, however, also had a significant portion of mapping to *Nona33virus*. Intriguingly, *Orthopoxvirus* was a dominant member of the virome comprising 19% of viral reads on average. Patients with STEC infections had 6% of reads belonging to *Orthopoxvirus*, though this percentage was not significantly different when compared to patients with the other three infection types (Mann Whitney U test p = 0.4875) or other infection types (Kruskal–Wallis p = 0.3671) (Figure 2.8A).

Differences in bacterial genera were also identified among cases when stratified by infection type (Figure 2.8B). *Bacteroides*, for instance, was a dominant bacterial member across samples from patients with all four types of enteric infections comprising an average of 42% of all reads (Figure 2.8). Specific bacterial populations, however, were also found to be more abundant in cases infected with specific pathogens. Genus *Salmonella* accounted for 20% of reads on average in *Salmonella* infections, which was significantly greater than the abundance in non-*Salmonella* infections (0.31%) (Mann Whitney U test p = 2.482e-09). ANCOM also identified *Salmonella* to be differentially abundant in samples from *Salmonella* cases relative to all other cases (Figure 2.8B). Genus *Shigella*, which comprised 6% of reads on average in the *Shigella* cases, was also significantly more abundant (Mann Whitney U test p = 0.0009327) compared to non-*Shigella* cases (1.5%); this difference was also confirmed using ANCOM. The bacterial reads in *Campylobacter* cases were comprised of 2% of *Campylobacter* on average and were significantly more abundant (Mann Whitney U test p = 8.197e-06) than the non-*Campylobacter*

cases on average (0.003%). Finally, *Escherichia* comprised 6% of reads on average in STEC infections, which was proportionally lower compared to non-STEC infections on average (10%), but this difference was not statistically significant when comparing STEC to non-STEC infections (Mann Whitney U test p = 0.88) or across infection types (Kruskal–Wallis p = 0.101), (Figure 2.8B). Interestingly, patients with Shigella infections had the greatest abundance of *Escherichia*; this percentage was significantly greater than the abundance in patients with STEC infections.

Hierarchical clustering identifies distinct fecal microbiome profiles

Four distinct clusters of microbiome profiles were identified. Cluster 1 (n=50) consists of 48% males (n=24) and 52% females (n=26). Cluster 2 (n=29) has 58.6% males (n=17) and 51.4% (n=12) females. Cluster 3 (n=44) has 69.8% males (n=30) and 30.2% (n=13) females. Cluster 4 (n=81) consists of 42.5% males (n=34) and 57.5% (n=46) females, (Table 2.4.). Additionally, the 19-64 age group is the most common age group across all 4 clusters; Cluster 1 (n=24, 48%), Cluster 2 (n=12, 41.4%), Cluster 3 (n=22, 51.1%), Cluster 4 (n=39, 38.7%). The second most common age group is 0-9 across all 4 clusters; Cluster 1 (n=12, 24%), Cluster 2 (n=8, 27.6%), Cluster 3 (n=13, 30.2%), Cluster 4 (n=21, 26.3%) (Table 2.4.).

Clusters also varied by health status. Cluster 1 (n=50) consists of 74% cases (n=37) and 26% controls (n=13), while Cluster 2 (n=29) is comprised 96.6% cases (n=28). Together (Clusters 1 and 2) accounted for 82% (n=65) of the cases (n=79). Cluster 3 (n=44) is 2.3% cases (n=1) and 97.7% controls (n=43). Cluster 4 (n=81) consists of 16% cases (n=13) and 84% (n=68) controls. Together (Clusters 3 and 4) account for 88% (n=111) of the controls (n=125), (Table 2.4.). Together (Clusters 1 and 2) accounted for 82% (n=65) of the cases (n=79). Cluster 3 (n=44) is 2.3% cases (n=1) and 97.7% controls (n=43). Cluster 4 (n=81) consists of 16% cases (n=13) and 84% (n=68) controls, (Table 2.4.). Together (Clusters 3 and 4) account for 88% (n=111) of the

controls (n=125) (Table 2.4.). Case hospitalization rates were consistent across clusters that had at least 10% of samples in the cluster as a case Cluster 1 (n=12, 35.1%), Cluster 2 (n=11, 39.3%), Cluster 4 (n=5, 41.7%). Diarrhea was reported as the most common symptom for all clusters that contained at least 10% of samples as cases, Cluster 1 (n=36, 97.3%), Cluster 2 (n=26, 100%), Cluster 4 (n=11, 84.6%). Bloody stool was reported in Cluster 1 (n=10, 27%) less than Cluster 2 (n=15, 85.2%). Bloody Diarrhea was most commonly reported in Cluster 2 (n=15, 57.7%). Fever was also frequently reported Cluster 1 (n=20, 62.5%), Cluster 2 (n=19, 79.2%), Cluster 4 (n=8, 72.2%).

PCA based on distances between microbiome samples showed that Clusters 3 and 4 are dominated mainly by non-infected samples and are localized within the right side of the PCA, while Clusters 1 and 2 are mainly dominated with cases and are localized more distally on the left side of the PCA (Figure 2.9A), and are distinct (NPMANOVA p = 0.006). Notably, Cluster 2 (orange) comprises 96.6% of cases and is the most distant and most heterogeneous cluster (Figure 2.9B). The differences between clusters can be quantified using standard diversity metrics.

Shannon index was significantly different across clusters (Kruskal–Wallis p = 1.193E-05) with case-dominated clusters (Cluster 1 and Cluster 2) having), for instance, had a lower Shannon diversity compared to Cluster index than Clusters 3 and 4. Cluster 4, which had the highest number of non-infected communities, had the highest level of diversity relative to all other clusters. By contrast, Cluster 2, comprising 96.6% of all cases, had the lowest diversity (Figure 2.10A), further supporting the differences identified between infected and non-infected communities (Figure 2.6). Richness was different across clusters (Kruskal–Wallis p = 2.2E-16). Case-dominated clusters (Clusters 1 and Cluster 2) had lower richness compared to Clusters 3 and 4 (Figure 2.10B). Evenness was significantly different (Kruskal–Wallis p = 0.006705) across clusters (Figure

2.10C). Additional analysis by case-status identified that these trends extend beyond cluster status. In general, Case communities found within Cluster 3 and Cluster 4 had higher diversity compared to cases communities belonging to Clusters 1 or 2. This trend also held up for non-infected communities, as those communities that clustered outside of Cluster 4 or were found outside Cluster 4 exhibited lower diversity comparable to that of the case samples. To investigate clinical impacts of these observations, we sought to identify disease associations with case-dominated clusters.

Gastroenteritis symptoms vary by microbiome composition

A univariate analysis was performed to identify differences in clinical outcomes among cases with microbiome profiles belonging to each case-dominated cluster, Clusters 1 or 2. Clinical-related characteristics (e.g., symptoms, hospitalization status) were classified as the exposure (independent variable), and Cluster designation (Cluster 1 or Cluster 2) was the outcome (dependent variable). Because Cluster 1 is more localized on the right side of the PCA (Figure 2.9A) towards the non-infected clusters, we hypothesized that illness would be less severe (e.g., no bloody diarrhea, chills, fever) with more non-specific symptoms (e.g., abdominal pain, nausea, fatigue). By contrast, Cluster 2 is the most distant cluster on the PCA (Figure 2.9A) from the non-infected communities, and hence, we hypothesized that Cluster 2 would have associations with more severe disease indicators such as bloody diarrhea, fever, chills, and vomiting. We, therefore, examined the distribution of symptoms reported by all 79 cases within each Cluster. In this analysis, Cluster 1 was associated with body aches (OR: 4.3, CI (95%): 1.5, 12.8) (Table 2.5.) and bloody diarrhea (OR: 0.4, CI (95%): 0.2, 1.1), whereas Cluster 2 was associated with vomiting (OR: 2.6, CI (95%): 1, 7.1) (Table 2.6.), bloody diarrhea (OR: 3.6, CI (95%): 1.3, 9.7). A strong association was also observed between Cluster 2 and headache (OR: 2.5, CI (95%): 0.8, 7.3),

though the difference was not statistically significant. We then created a severe disease index score that consisted of hospitalization or a history of bloody diarrhea and fever as fever, and bloody diarrhea was significantly associated with each other (p=0.02). Importantly, patients with microbiome profiles belonging to Cluster 2 (n=28; 34.5%) were significantly more likely to have a higher severe disease index score when compared to patients with profiles belonging to all other Clusters (n=51; 65.5%). No associations were identified with sex, age, or race for either Cluster 1 (Table 2.5.) or Cluster 2 (Table 2.6.). Based on these different associations with disease, we decided to investigate the differences in the microbiome between the clusters.

Specific viral and bacterial populations dominate in case clusters

Given the differences in the types of symptoms reported by cases with profiles belonging to Cluster 2, we sought to identify specific viral and bacterial populations that define each Cluster. Cluster 1 (green), for instance, has a microbiome composition that is more like that of Clusters 3 and 4 with minor alterations (Figure 2.11). Cluster 2 (orange), which was associated with more severe clinical symptoms, had the most distinct microbiome compared to the other three Clusters. It was hypothesized that Cluster 1 and Cluster 2 would have a common core microbiome since both are comprised mainly of cases with infections. We further hypothesized that Cluster 2 communities would have a distinct profile from Cluster 1 communities because of differences noted in symptom profiles of the patients.

In total, 17 genera were shared across the four clusters which consisted mainly of Proteobacteria, Bacteroides, and Firmicutes. Additionally, 52 genera were unique to Cluster 2 from all other clusters (Figure 2.12A). In case clusters, ANCOM identified 24 differentially abundant genera in Cluster 1 and 86 in Cluster 2; 18 of the taxa identified were shared between communities representing both Clusters (Figure 2.12B). *Orthopoxvirus* was present in all four clusters but was

highest in abundance in Cluster 2 and Cluster 1 (Figure 2.12C). *Nona33virus* was highest in abundance in Cluster 2. Bacteroidetes was detected in all 4 clusters at varying abundance (Figure 2.12D). Proteobacteria (*Escherichia*, *Salmonella*) were highest in Cluster 2, and *Escherichia* was second highest in Cluster 1. *Alistipes* were detected highest in Clusters 3 and 4, with *Prevotella* being both differentially abundant by ANCOM and highest in abundance in Cluster 3 (Figure 2.12D). Further analysis of the case clusters identified specific changes.

The common microbiome shared between Cluster 1 and Cluster 2 communities are dominated by Proteobacteria and include genera representing common enteric pathogens (e.g., *Salmonella, Escherichia, Shigella*) as well as Bacterioidetes (*Alistipes*), Firmicutes (e.g., *Oscillibacter, Neglecta*) and Orthopoxviris (Table 2.7.). The distinct microbiome of Cluster 1 includes six genera representing Bacteriodetes (*Odoribacter*), Firmicutes (*Christenselnella*), and Others (*Lachnoclostridium, Akkermansia Veillonella*, *Asaccharobacter*). Cluster 2, however, is defined by 68 additional genera that are not found in Cluster 1 communities. Viruses represent 36.7% of this difference (n=25 genera), and 96% of these viral taxa belong to Caudovirales and include Podoviridae (n=5 genera), Siphoviridae (n=12 genera), Myoviridae (n=8) and one eukaryotic virus (*Cytomegalovirus*) (Table 2.5.).

Bacteria genera (n=46) that are differentially abundant in Cluster 2 communities relative to the rest of the samples consisted of 65% Firmicutes (n=30 genera), 11% Bacteroidetes (n=5 genera), 11% Proteobacteria (n=5 genera), and 13% Others (n=6 genera). Network analysis of the differentially abundant taxa in Cluster 2 demonstrates that many are strongly correlated with one another (Figure 2.13). Enterobacterales (red) were positively correlated (green edges) with other pathogenic bacteria (Figure 2.13), including Lactobascillales (*Enterococcus* and *Streptococcus*), Bascillales (*Staphylococcus*), and Psuedomonindales (*Acinetobacter*, *Psudomondonas*).

Enterobacterales were also strongly correlated with viruses belonging to Orthopoxvirus and Caudovirales (*P2virus*, *Nona33virus*, *P22virus*). Additionally, Enterobacterales were primarily negatively correlated (red edges) with Clostridales (green) and Bacteroidales (yellow), (Figure 2.13), while bacteriophage (pink) was negatively correlated with Clostridales (green).

A univariate analysis was performed to identify taxa associated with Cluster 2 (Table 2.8.). Taxa were selected based on ANCOM analysis, and samples were evaluated for the relative abundance of a given Genus that was above or below the normalized average in order to identify the taxa that were higher in abundance for Cluster 2. Taxa were the exposure (independent variable), and Cluster 2 designation was the outcome (dependent variable). Cluster 2 was found to be associated with the following virus genera that were above the study average; *Orthopoxvirus* (OR: 19.4, CI (95%): 7.7, 48.9) and *Cytomegalovirus* (OR: 3.0, CI (95%) 1,8.6), common enteric bacteriophage, *Nona33virus* (OR: 17, CI (95%): 6.1, 47), *P22virus* (OR: 6.6, CI (95%): 2.2, 20.1), *P2virus* (OR: 9.4, CI (95%): 3.3, 27.1), and *P1virus* (OR: 7.3, CI (95%): 2.2, 24.7). Additionally, associations were identified with bacteria genera that were above average in the study. Common enteric pathogens such as *Salmonella* (OR: 9.7, CI (95%): 3.5, 26.9), *Escherichia* (OR: 12.6, CI (95%): 5.2, 30.3), and *Shigella* (OR: 19.4, CI (95%): 7.7, 48.9) were highly abundant in Cluster 2 communities relative to other Clusters. Other pathogenic bacteria such as *Enterobacteria* (OR: 14.1, CI (95%): 5.7, 34.6), *Pseudomonas* (OR: 10.9, CI (95%): 2.4, 56.7), *Staphylococcus* (OR: 12.1, CI (95%): 4.7, 31.4), *Haemophilus* (OR: 9.1, CI (95%): 3.5, 24), *Acinetobacter* (OR: 19.5, CI (95%): 7, 53.9), *Streptococcus* (OR: 10.1, CI (95%): 3.8, 27) and those that classify as opportunistic pathogens like *Serratia* (OR: 25.7, CI (95%): 9, 73.5) were also highly abundant (Table 2.8.).

Logistic Regression for predicting Cluster 2 status

Based on the different associations with Cluster 2 status as well as the network analysis (Figure 2.13), we sought to build a model that incorporated these associations. The base model was built while including the presence of genera representing common enteric pathogens, namely *Salmonella*, *Escherichia*, and *Shigella*. The inclusion of *Salmonella* was determined to have the most substantial effect on the base model. Bacteriophages that infect Enterobacteriaceae (*P22virus*, *P2virus*, *Nona33virus*, *Lambdavirus*) were subsequently incorporated into the model (Table 2.7.) but were removed as more significant variables could explain their incorporation such as *Salmonella*. The Hosmer-Lemeshow goodness-of-fit test was evaluated to determine if the model was being overfitted. Wald's test was used to incorporate significant variables. The final model demonstrated that *Acinetobacter*, *Orthopoxvirus*, *Salmonella*, and *Serratia* were the critical predictors of Cluster 2 communities (Table 2.8.).

DISCUSSION

Studying the microbiome has traditionally utilized amplicon sequencing of targeted genes (16S) or shotgun sequencing. A recent study of 49 samples, which represents one of the most extensive paired comparisons to date, examined differences between shotgun metagenomics and amplicon sequencing. This study found that there were significant differences between the reported biodiversity for both methods (48). In brief, the shotgun metagenomics reported less richness overall but agreed consistently with amplicon sequencing in the taxonomy that was reported. This finding seems to be at odds with the other studies that have performed amplicon sequencing and found an increased richness with Illumina sequencing (49, 50). Small sample size studies ($n < 10$) that examined shotgun sequencing using the Illumina platform found increased richness and a better representation of community structure (51, 52). The human microbiome project performed marker gene analysis with shotgun sequencing (53) and compared it to amplicon sequencing ($n=51$). While it was concluded that the use of shotgun sequencing is a better approach, the reported results did not show significant differences at the genera level between shotgun sequencing and amplicon sequencing. Comparing across studies can be difficult with microbiome studies as there is not a consensus in the field on how comparable one set of results is to another, and the results can be reflective of the platform and approach being utilized (52). Studies can be compared by utilizing significant findings, even though differences in the study design might differ.

The amount of bacteria reads in a fecal metagenome typically comprises >90% of reads (54), whereas viruses have been reported from 5.8% (27) to 22% (55) of the fecal microbiome, our identification rate was 0.5% on average across samples. We were able to achieve an annotation rate of 90% for reads at the Phylum-level. However, it is difficult to discern if a bacteriophage is

either a prophage or extracellular based on metagenomes, which could, in part, explain our lower identification rate of viruses. Additionally, our average coverage for all samples is 78%, which is lower than the ideal (>95% is saturation). A previous study found that increasing the sequencing depth by 2x resulted in a 3.3% increase in the number of genera present, with most of the increases attributable to rarer taxa like bacteriophage (56). There are, however, several significant limitations to the previous study, which cast doubt on its value. The sample size (n=8) was low and could lead to a spurious association. Both the Shannon and inverse Simpson indices were not associated with increased sequencing depth. The Shannon index represents the richness over the evenness in the community; hence, an increased richness should accompany changes in the Shannon index. Also, there was no mention of post-hoc corrections (Benjamin-Hochenberg) to control for false discovery in pairwise comparisons, which could lead to a false association due to a higher type I error rate. The Nemenyi post-hoc comparison utilized (56) is inappropriate for microbiome datasets (57). Indeed, as sequencing depth is increased, there is a higher chance of sequencing regions of the (meta) genome that can be annotated to a lower taxonomical level. However, we did not find that increasing sequencing depth was associated with increased richness in our study, which has much higher statistical power (n=204). Our depth and coverage are lower comparatively (56), so we might not have achieved enough sequencing depth to uncover the association reported with sequencing depth and richness (56). Despite these limitations we were still able to find trends in the analysis.

The findings here on the microbiome are in line with our previous findings with 16S amplification (11). In total, we examined 204 fecal microbiomes (79 cases, 125 healthy controls) using metagenomics. The richness and Shannon index were both significantly higher in control (healthy groups) relative to the cases in this study and previously (11). We estimated 109 to 173

OTUs to be present in our amplicon analysis (11). Herein, we detected greater richness comparatively at the genera level with 2,659 genera; interestingly, 150 genera collectively accounted for 99% of our annotated reads, which is similar to the total number of OTUs in the previous study (11). We initially observed that case samples contained a higher proportion of reads on average that were annotated as human (15.2%) compared to controls (0.1%). The increased presence of human DNA in stool samples is a component of dysbiosis in *Clostridium Difficile* infection (58), IBD (59), and colorectal cancer (60), as such the increased presence of reads annotated as human in this study could be due to inflammation-induced tissue destruction present in cases from hemorrhagic colitis. The tissue destruction could lead to release of nutrients for the microbiota including carbon sources, vitamins, minerals like iron. Iron is necessary for the growth of many different strains of bacteria. Iron-acquisition by some bacteria has been linked to more invasive phenotypes (61) and is tightly regulated by the human body with siderophores. The release of the cellular contents could provide the necessary nutrients to drive the observed dysbiosis in gastroenteritis. A future investigation into the metabolic profiles of the reads present in this study would likely yield enrichment in iron-scavenging pathways within cases.

The main finding in cases of gastroenteritis is increased Proteobacteria (10, 11). Proteobacteria, a dominant phylum, is associated with inflammation and is a signature of dysbiosis in many disease states including gastroenteritis (10, 62–65). Cases had a higher abundance of Proteobacteria compared to uninfected controls. Uninfected controls had higher abundances of Bacteroidetes and Firmicutes compared to cases. Both of these findings have been observed previously (11) and in literature (10, 66). These findings continue to the Family-level where Enterobacteriaceae was dominant in cases that likely represents either an increase in *Escherichia* abundance or the pathogen. Differential analysis at the genera level previously identified that

Roseburia, *Blauta,* and *Lachnospiraceae* were most differentially abundant in healthy people, (11) which we can affirm was differentially abundant in our analysis with ANCOM. We also identified that decreased relative abundance in *Roseburia* was associated with more severe illness.

The dominant virus order detected in our study was Caudovirales, with Siphoviridae being the most abundant family of virus, which has been reported previously (67). Microviridae is also commonly found in healthy populations, and increases with age (68, 69), which we also observed was increased in abundance in our healthy controls. We also noted correlations between infection types and the microbiome as noted previously (11), but we expanded on those findings here. *Salmonella* infections had increased proportions of *P22virus* which is a genus of mostly *Salmonella* prophages (70). *Nona33virus* is a recently recognized genus, which consists of *stx*-harboring bacteriophage that infect *Escherichia* (71) and was most abundant in STEC infections. *P1virus* were found to be specific for Shigella infections. Pathogen genomes harbor many prophages and we would expect to detect these at about equal frequency if these prophages were not active. Since the detected prophage listed above are differentially abundant, we expect that these are actively replicating phage and are most likely lytic.

Phage may control populations of common commensal bacteria, like *Enterococcus faecalis.* A phage in *E. faecalis* can integrate into two distinct regions in the host genome (72). Expression of each insertion is regulated by nutrient availability, and, in optimal growth conditions for *E. faecalis,* the phage switches from the lysogenic phase to lytic, which prevents over-expansion of the niche used by *E. faecalis* in the gut (72). This example highlights the "kill-the-winner" (KTW) dynamic (73), which represents an expansion of the predator-prey Lotka-Volterra model (74). The KTW model is also applicable to other ecological systems and was initially defined based on observations of the ocean microbiome (75). KTW dynamics predict the

expansion of bacterial population results in a corresponding increase in the phage population that maintains individual populations, thereby increasing overall stability (76). The oral microbiome of five healthy individuals contained mostly lysogenic phages, suggesting these viruses may have a similar role in shaping this microbiome (77). The Bacteriophage Adhering to Mucus (BAM) model proposes phage localize and adhere to mucous membranes in the host. Cell culture work performed *in vitro* found that mucus-producing human colon epithelium is more protected from bacterial invasion if combined with a phage inoculum (78). Taken together, the BAM, KTW, and Lotka-Volterra models propose that phage regulate bacterial populations (79) and can actively ward off pathogens (78). Our findings support these models in that the presence of Enterobacteria-phage being present alongside increases in their host Enterobacteriaceae.

Enterotypes (27) are groupings or clusters of samples that have a typical microbial composition. Enterotypes exist independent of age and gender but can be influenced by diet (27), and their value is debated (80). We further grouped samples by total microbiome composition, as we did previously (11). Clustering offers immense benefit to microbiome research as it allows grouping of samples by similarity (or dissimilarity) with complex datasets. Studies on gastroenteritis have identified a subset of patients that will exhibit a shift to an *E. coli-Shigella* dominated enterotype, which is independent of the infecting agent (10). We can confirm that our dataset matches previous findings (10, 11). Cluster 1 and Cluster 2 which consisted of cases were Enterobacteriaceae dominant regardless of infection-type. Additionally, Clusters 3 and 4, which consisted mostly of samples from the healthy controls, were comprised mostly of *Prevotella* and *Bacteroidetes*, respectively, and could represent observed enterotypes based on diet (27). We did not directly assess enterotype status because of limitations in dietary information, but the findings here show the clustering of our controls differentiated by dominant enterotype-specific taxa.

Our previous study of an overlapping subset of 275 samples identified five clusters using 16S rRNA sequencing (11). In this study of 204 samples, we identified four clusters that differed significantly from each other. While the inclusion of additional samples could have boosted the statistical power and might have split the dataset into two case and three control clusters as we found previously (11), similar results were observed across clusters. The majority of cases were found in Cluster 1 and Cluster 2 and were more likely to report symptoms, as reported previously (11). Similarly, we found that patients in Cluster 2 were at risk for more severe illness than other patients and had significant alterations in their microbiome which our previous study did not identify (11). These findings are due in part to the increased richness that can be assessed with shotgun sequencing. Cluster 1 was associated with more mild symptoms (body aches) and had a microbiome profile more like the controls. Similar associations have been reported that there is less severe disease in patients that have microbiome profiles more similar to uninfected controls (81). Cluster 1 was found to be associated with *Veillonella* which degrades bile acid (10) and has been associated with gastroenteritis previously (10). Cluster 1 was also associated with *Akkermansia* which degrades mucin (82), resulting in mucosal degradation (83). The degradation of mucin could directly release bacteriophage localized within the mucosa and subsequently infect nearby bacteria. Such disruptions will undoubtedly impact the microbiota composition and alter immune system responses; further investigation is warranted. *Odoribacter* was also associated with Cluster 1 and produces many small-chain fatty acids including butyrate, acetate, and propionate (84), disruption of which alters inflammation. Other taxa identified but of unknown significance include *Lachnoclostridium,* which has been linked to colorectal cancer (85) and possibly present due to the inflammation, and *Christensenella,* which evidence suggests might be a keystone species for the healthy microbiome (86). Previous research has also correlated

abdominal pain with *Alistipes* (65) which was common to both Cluster 1 and Cluster 2 and *Staphylococcus* (81) which were specific to Cluster 2.

Additionally, Cluster 2 was associated with more severe symptoms (bloody diarrhea and vomiting) and had more severe dysbiosis. Differentially abundant taxa identified specific to Cluster 2 include *Roseburia*, which was decreased in Cluster 2 and could create a pro-inflammatory environment since *Roseburia* produces butyrate (87) that has been shown to decrease inflammation (88). *Acinetobacter* was also elevated and may play a role in an immune response. Importantly, recent evidence suggests that it can directly cause differentiation of T cells *in vitro* but also downregulates helper T cells (89), potentially altering the response by the immune system to the dysbiosis. *Enterococcus*, a common commensal, was also found to be elevated. Because *Enterococcus* has been shown to produce bacteriocins that have strong antimicrobial properties (90), the microbial population could have an impact on the growth and survival of other bacteria.

Cluster 2 was also associated with many changes in viral composition, most of which directly utilize Enterobacteriaceae as the host. Caudovirales were increased for both gastroenteritis patients and within Cluster 2 communities; similar findings were seen in a study on IBD. Specifically, phage increased in abundance and diversity within IBD patients, while the bacterial population was conversely decreased (20), and blooms in bacteriophage have been tied to increases in host inflammation (91) and were found to affect the bacterial population directly (92). Expansions in Caudovirales have also been noted in viromes of immunocompromised HIV-infected patients, who have altered pro-inflammatory microbiomes and increased *Adenovirus* abundance compared to healthy populations (63).

Eukaryotic viruses can affect the host immune system, as well. *Orthopoxvirus*, for instance, produces soluble molecules that bind chemokines, cytokines, and interferon to alter the immune response (93, 94). Testing in mouse models has elicited distinct microbial profile changes, which included decreases in Proteobacteria compared to mock (95). Other mouse models have also shown the importance of eukaryotic viruses. The presence of murine norovirus in germ-free mouse models restored the typical morphology of the intestinal tract through a signaling cascade without an overt immune response to the virus (96), suggesting that eukaryotic viruses can support the restoration of bowel homeostasis. Another study concluded that the presence of inactivated rotavirus could reduce inflammation in the colon through induction of anti-inflammatory cytokines acting on toll-like receptors (97). Additional confirmation is needed to confirm the findings of *Orthopoxvirus*, as this finding has been determined to be a false positive in other studies (63). However, in contrast to this study, these studies utilized viral only databases with BLAST with a standard e-value (10^5) given the smaller database size of the viral only databases a smaller e-value should be utilized. Here, we used the totality of the NCBI non-redundant database with a kmer-based approach. At minimum, the identification of a sequence as *Orthopoxvirus* had to have a higher score compared to all other non-viral signatures in the database. Nonetheless, additional analysis is needed to confirm the presence of *Orthopoxvirus* and if it does indeed taper the immune system during acute gastroenteritis. Culturing would be ideal for confirmation of the findings presented, though many of the taxa identified are non-cultivable.

Gastroenteritis can have two types of effects on human health — an acute effect results from a which involves immediate pathogen infection.For instance, *Salmonella* can directly exploit inflammation to colonize the GI tract resulting in diarrheal illnesses (98) and increased abundance of Enterobacteriaceae (99) which ultimately resolves. We observed findings related both to a

Salmonella infection resulting in increased *Enterobacteriaceae*. If the microbiome alteration does not resolve, a chronic inflammatory state can develop; with symptoms lasting up to 10 years after incident (100, 101) or inflammatory bowel disease (IBD). The year following a case of infectious gastroenteritis, individuals are 2.4 times more likely to develop IBD (102). An underlying mechanism towards the chronic state has been proposed. The pathogen initiates an inflammatory state-driven primarily by host immunity (99, 103), this creates an environment for a pathobiont, a resident microbe that has pathogenic potential, such as adherent-invasive *Escherichia coli* (AIEC) to bloom (104), which we have identified *Escherichia* here and previously (11) as being increased in abundance in cases compared to control populations regardless of infecting agent. Sensitization of the host defense to AIEC in mouse models prevented the bloom that occurred and improved health in murine models (104). Additional research is needed to determine if *Escherichia* can be prevented from blooming during acute gastroenteritis in humans. Examining the effects of iron on the microbiota may reveal a potential therapy, as an intervention could examine iron effects on the microbiota, specifically *E. coli*, to determine if iron-chelating agents could prevent further dysbiosis in gastroenteritis. The development of such a therapeutic would lower the disease burden of gastroenteritis and could potentially lower the incidence of chronic sequelae related to gastroenteritis such as IBS and IBD.

In short, we aimed to analyze both the viral and bacterial signatures simultaneously in acute bacterial gastroenteritis. Cluster 2 had a more substantial proportion of viral reads present, which could be due to high rates of bacteriophage induction in response to changes in specific bacterial host populations. Additionally, the logistic regression model identified a common enteric pathogen (*Salmonella*), an opportunistic pathogen (*Serratia*), a bacterium that directly interacts with the immune system (*Acinetobacter*) and a eukaryotic virus (*Orthopoxvirus*), which also directly

interacts with the immune system but potentially opposes *Acinetobacter* to be the critical predictors of Cluster 2 communities. Although the study is limited by sample size (n=204) and sequencing (average coverage = 78%), cross-assembly (55) of the *Orthopoxvirus* sequences in this study could further validate the findings by achieving a more specific signature for annotation. Cross-assembly is a computationally intensive process and would require substantial resources to complete but would add considerable value as a follow-up study. Assemblies were not directly utilized in this study because of the statistical differences noted in mapping frequencies between cases and controls. In a future study, assemblies could provide a more specific signature for annotation. Additional studies are needed that directly assess the RNA virome, which remains an overlooked component of virome studies in general. Direct isolation of viruses, in combination with sequencing, is recommended, as studies of the virome remain primarily limited by lack of known viruses.

APPENDIX

Table 2.1. Sequencing quality and coverage estimates for 204 metagenomes Results for the total sequencing (column 2), the quality control (columns 3-4), annotation results (column 5) and overall coverage (column 6).

Study ID	Reads Paired-forward total Count (Gbp)	Reads remaining after low-quality read removal Count (%)	Reads remaining after human read removal Count (%)	Reads annotated Total (%) Viral (%)	Nonpareil Coverage (%)
ER0043	3690286 (1.1)	2551094 (69.1)	2550606 (100)	69.815823 (0.01)	87.1
ER0073	4475020 (1.3)	2074677 (46.4)	1567523 (75.6)	68.416818 (0.07)	82.6
ER0087	3773902 (1.1)	2545292 (67.4)	1933497 (76)	79.21824 (42.81)	77.6
ER0109	3581466 (1.1)	2449737 (68.4)	2356690 (96.2)	63.66909 (0.01)	84
ER0114	4322314 (1.3)	3059178 (70.8)	3059085 (100)	66.012962 (1.08)	83.5
ER0117	4278113 (1.3)	2978060 (69.6)	159315 (5.3)	38.54845 (0.01)	27.2
ER0130	2462439 (0.7)	1694157 (68.8)	1694025 (100)	60.618575 (0.04)	65.4
ER0151	388095 (0.1)	231689 (59.7)	230122 (99.3)	65.682525 (1.57)	53.3
ER0152	232570 (0.1)	125615 (54)	125096 (99.6)	66.652544 (0.01)	26.1
ER0163	586349 (0.2)	364476 (62.2)	363875 (99.8)	51.344703 (0.08)	45
ER0189	2237322 (0.7)	1579033 (70.6)	1578907 (100)	60.130712 (1.91)	78.8
ER0190	1944382 (0.6)	1354015 (69.6)	1352769 (99.9)	70.613134 (0.01)	74.9
ER0191	2156229 (0.6)	1345831 (62.4)	1345436 (100)	63.110984 (0.01)	70.3
ER0192	545806 (0.2)	307626 (56.4)	307007 (99.8)	65.777413 (0.01)	44.3
ER0194	1837552 (0.6)	1422110 (77.4)	111763 (7.9)	39.218257 (0.03)	34.1
ER0196	4906306 (1.5)	3222852 (65.7)	3208766 (99.6)	47.144177 (0)	94.2
ER0201	1276119 (0.4)	871703 (68.3)	842575 (96.7)	67.516847 (0.01)	70.3
ER0203	2912425 (0.9)	2109330 (72.4)	2035748 (96.5)	59.215179 (0.01)	71.5
ER0206	2802969 (0.8)	1965133 (70.1)	1746050 (88.9)	51.707844 (0.01)	79.6
ER0210	444000 (1.3)	311323 (70.1)	310743 (99.8)	69.059241 (0.01)	37.4
ER0222	2667880 (0.8)	1822512 (68.3)	1822313 (100)	57.56992 (0.06)	69.9
ER0224	5897349 (1.8)	4028915 (68.3)	4028656 (100)	57.281524 (0)	90.2
ER0225	4323343 (1.3)	2885169 (66.7)	2885122 (100)	52.604623 (1.24)	80.4

Table 2.1. (cont'd)

ER0226	379000 (1.1)	263194 (69.4)	263167 (100)	60.683752 (0.01)	48.3
ER0228	5384311 (1.6)	3651547 (67.8)	3651393 (100)	71.260326 (0.02)	79.1
ER0229	5268224 (1.6)	3555546 (67.5)	3554925 (100)	76.039992 (0.01)	88.4
ER0230	2110424 (0.6)	1423770 (67.5)	1423491 (100)	56.915487 (0.2)	56.2
ER0231	5123000 (1.7)	3648988 (71.2)	3648390 (100)	50.212805 (0.01)	83.5
ER0236	2844829 (0.9)	1890144 (66.4)	1889704 (100)	71.135879 (0.03)	79.5
ER0237	544495 (0.2)	306045 (56.2)	304761 (99.6)	56.536419 (0.02)	41.8
ER0238	4707792 (1.4)	3295554 (70)	3295476 (100)	59.955817 (0.03)	61.8
ER0240	1763312 (0.5)	1215983 (69)	1215722 (100)	55.565508 (0.49)	45.6
ER0241	1426161 (0.4)	924030 (64.8)	923599 (100)	51.784065 (0.96)	65.9
ER0242	4328121 (1.3)	3033114 (70.1)	3032927 (100)	45.141548 (0.94)	83.8
ER0243	3087983 (0.9)	2280010 (73.8)	2279662 (100)	53.469698 (0.11)	64.8
ER0244	4423459 (1.3)	3054924 (69.1)	3054767 (100)	68.105447 (0.49)	83.9
ER0245	4001677 (1.2)	2733059 (68.3)	2732743 (100)	59.360523 (0.47)	68.8
ER0246	4907257 (1.5)	3382923 (68.9)	1913367 (56.6)	52.568997 (0.04)	67.1
ER0273	3910156 (1.2)	2682649 (68.6)	2681579 (100)	66.538159 (0.01)	88.2
ER0289	4104095 (1.2)	2792814 (68)	2792161 (100)	59.057322 (0.02)	83.3
ER0290	182295 (0.1)	112589 (61.8)	112407 (99.8)	46.859599 (0.9)	37.1
ER0291	4336392 (1.3)	2848156 (65.7)	2847519 (100)	68.727434 (0)	81.3
ER0301	3125782 (0.9)	2105356 (67.4)	2092588 (99.4)	68.398843 (0)	80.5
ER0303	3015911 (0.9)	1890784 (62.7)	1890241 (100)	61.867799 (0.01)	78.4
ER0304	4194383 (1.3)	2812539 (67.1)	2812197 (100)	56.214372 (0.01)	78.3
ER0305	4261812 (1.3)	2804294 (65.8)	2803791 (100)	61.269584 (0.04)	84.9
ER0332	3238539 (1)	2174562 (67.1)	2161531 (99.4)	56.61745 (0.01)	75
ER0379	2523145 (0.8)	1673988 (66.3)	1673843 (100)	59.533014 (0.01)	52.4
ER0380	2526254 (0.8)	1682934 (66.6)	1682756 (100)	64.027065 (0.01)	66.5
ER0438	3876292 (1.2)	2212919 (57.1)	2212507 (100)	61.098235 (0.02)	63.6
ER0443	3654000 (1.6)	2514033 (68.8)	2513642 (100)	71.948559 (0.01)	77.4
ER0444	4647744 (1.4)	3230445 (69.5)	3230348 (100)	73.648233 (0.01)	84.4

Table 2.1. (cont'd)

ER0445	3527659 (1.1)	2404223 (68.2)	2403186 (100)	67.417172 (0.04)	79.9
ER0628	4143623 (1.2)	2860706 (69)	2737187 (95.7)	62.884808 (0)	82.3
ER0631	4947199 (1.5)	3491067 (70.6)	449651 (12.9)	66.886871 (1.71)	33
ER0640	2264000 (0.9)	1583527 (69.9)	1521089 (96.1)	74.422937 (0)	85.7
ER0641	1010000 (1.5)	709688 (70.3)	20171 (2.8)	47.109812 (0.01)	13.3
ER0644	42000 (1.3)	30525 (72.7)	30449 (99.8)	49.513924 (0.01)	10.4
ER0646	984000 (0.9)	700612 (71.2)	698633 (99.7)	61.67966 (0.01)	59
ER0649	3253859 (1)	2358811 (72.5)	2321899 (98.4)	59.499313 (0.64)	67.1
ER0653	5145388 (1.5)	3012888 (58.6)	1807740 (60)	58.371776 (0.02)	66.6
ER0661	4139304 (1.2)	2909838 (70.3)	2389877 (82.1)	56.432873 (0.02)	77.9
ER0676	1844000 (1.3)	1291819 (70.1)	200484 (15.5)	60.393479 (0.02)	23.6
ER0680	3452654 (1)	2491927 (72.2)	2484834 (99.7)	49.292036 (0.03)	92
ER0693	3819850 (1.1)	2601786 (68.1)	2580787 (99.2)	63.332222 (0.16)	82.2
ER0694	3842806 (1.2)	2815063 (73.3)	1925995 (68.4)	67.870331 (0)	67
ER0708	5468210 (1.6)	3536021 (64.7)	3535631 (100)	66.950614 (0)	91.9
ER0003	2503344 (1.3)	2416794 (96.5)	2416565 (100)	51.377703 (0.01)	84.7
ER0075	2513186 (1.3)	2443893 (97.2)	2437945 (99.8)	34.516915 (0)	99.2
ER0092	2341075 (1.2)	2181952 (93.2)	1638581 (75.1)	23.65799 (0)	79.4
ER0093	2438656 (1.2)	2339042 (95.9)	2338654 (100)	57.338745 (0.07)	70.3
ER0126	2887334 (1.4)	2770758 (96)	2770687 (100)	63.744739 (0.02)	81
ER0209	3463776 (1.7)	3199411 (92.4)	3195012 (99.9)	44.553532 (0.01)	94.4
ER0264	2712638 (1.4)	2416483 (89.1)	1893607 (78.4)	44.312638 (0)	75.8
ER0265	2018980 (1)	1942334 (96.2)	1930233 (99.4)	82.84611 (0)	96.4
ER0275	2763759 (1.4)	2688130 (97.3)	2590963 (96.4)	69.863941 (0.01)	84
ER0294	2302743 (1.2)	2231233 (96.9)	2230377 (100)	55.897398 (0.08)	78
ER0299	2914589 (1.5)	2826233 (97)	2787126 (98.6)	78.899513 (0.06)	91.8
ER0331	2756416 (1.4)	2683862 (97.4)	2683205 (100)	55.206048 (0.09)	95.1
ER0376	2949070 (1.5)	2860618 (97)	2843485 (99.4)	83.066788 (0)	95.6
ER0377	2982274 (1.5)	2880774 (96.6)	2843064 (98.7)	60.841207 (0.07)	80.2

Table 2.1. (cont'd)

ER0385	3268838 (1.6)	3152845 (96.5)	3151489 (100)	72.824214 (0.01)	80.2
ER0487	3032451 (1.5)	2937190 (96.9)	2937016 (100)	61.717564 (1.09)	89.9
ER0510	4011214 (2)	3881859 (96.8)	3778439 (97.3)	48.601768 (0.01)	92.5
ER0513	3647954 (1.8)	3259175 (89.3)	979802 (30.1)	66.217415 (0.01)	45.7
ER0518	3734475 (1.9)	3560493 (95.3)	1838195 (51.6)	69.665266 (0.01)	59
ER0519	2502866 (1.3)	2429050 (97.1)	2416803 (99.5)	66.073663 (0)	94.1
ER0522	3761551 (1.9)	3629116 (96.5)	3628367 (100)	77.551334 (0.02)	96.7
ER0535	2287524 (1.1)	2208850 (96.6)	2089682 (94.6)	63.897697 (0.02)	84.6
ER0556	2879420 (1.4)	2802796 (97.3)	2802205 (100)	60.706754 (0.02)	90.5
ER0557	3482958 (1.7)	3342524 (96)	2738584 (81.9)	65.538704 (0.07)	68.4
ER0562	3865705 (1.9)	3669968 (94.9)	1650057 (45)	58.308875 (0.01)	51.6
ER0563	2054040 (1)	1963311 (95.6)	1537503 (78.3)	46.312936 (0)	73
ER0567	2686650 (1.3)	2586383 (96.3)	2565135 (99.2)	69.992498 (0)	82.5
ER0568	2233653 (1.1)	2156875 (96.6)	2137755 (99.1)	72.081066 (0.01)	90.6
ER0569	2717762 (1.4)	2593582 (95.4)	2591240 (99.9)	49.520458 (0.03)	80.5
ER0576	2483636 (1.2)	2403597 (96.8)	2317317 (96.4)	53.035119 (0.01)	92.1
ER0599	1989731 (1.4)	1926928 (96.8)	1918450 (99.6)	63.709328 (0.6)	43.6
ER0610	2856204 (1.3)	2648770 (92.7)	942730 (35.6)	75.444685 (0.01)	82.5
ER0642	2659009 (1.3)	2575434 (96.9)	2567963 (99.7)	82.258479 (0)	94.5
ER0682	3396048 (1.7)	3270813 (96.3)	158886 (4.9)	40.450769 (0)	30.5
ER0702	2854421 (1.4)	2715943 (95.1)	2713974 (99.9)	74.457471 (2.83)	92.6
ER0730	4321568 (2.2)	4150575 (96)	4149231 (100)	64.124949 (0.11)	91.6
ER0751	2832541 (1.4)	2745009 (96.9)	2744922 (100)	38.388086 (0)	94
ER0769	2953626 (1.5)	2844938 (96.3)	2837734 (99.7)	61.962237 (0)	93.4
ER0775	2937369 (1.5)	2834883 (96.5)	2831084 (99.9)	37.274218 (0)	89.6
ER0776	3328875 (1.7)	3220930 (96.8)	3219376 (100)	40.818349 (0)	95.8
ER0785	2704803 (1.4)	2613298 (96.6)	2608021 (99.8)	59.258712 (0.01)	89.2
ER0794	2943331 (1.5)	2853100 (96.9)	2842644 (99.6)	61.341343 (0)	90.1
ER0831	2669794 (1.3)	2576760 (96.5)	2558749 (99.3)	22.638276 (0)	95.7

Table 2.1. (cont'd)

ER0853	2321135 (1.2)	2103992 (90.6)	1897677 (90.2)	47.315647 (0)	87.5
ER0859	2507476 (1.3)	2430524 (96.9)	2428609 (99.9)	70.452843 (0)	94.3
ER0868	2377090 (1.2)	2311151 (97.2)	2310675 (100)	75.008731 (0.01)	89.6
ER0902	2953083 (1.5)	2847916 (96.4)	523324 (18.4)	42.10598 (0.01)	36.9
ER0129	5858632 (2.9)	5688595 (97.1)	5688477 (100)	57.962686 (0.82)	90
ER0188	3049908 (1.5)	2977225 (97.6)	2977173 (100)	69.675386 (0.03)	88.9
ER0217	501000 (1.3)	474829 (94.8)	474780 (100)	49.220898 (0.01)	77.4
ER0218	2191259 (1.1)	1896424 (86.5)	1895129 (99.9)	60.875956 (0.01)	72.7
ER0219	2976614 (1.5)	2862329 (96.2)	2861903 (100)	82.163795 (0.01)	85
ER0220	3251021 (1.6)	2687907 (82.7)	2685636 (99.9)	65.072276 (0.01)	81.5
ER0223	3035510 (1.5)	2957218 (97.4)	2957211 (100)	68.798462 (0.03)	85.3
ER0249	2786305 (1.4)	2534620 (91)	2532467 (99.9)	57.504394 (0.03)	78.7
ER0250	2842178 (1.4)	2505249 (88.1)	2505155 (100)	64.732807 (0.92)	87.1
ER0256	3377998 (1.7)	3297891 (97.6)	3297811 (100)	64.527671 (0)	76.4
ER0257	2731396 (1.4)	2669738 (97.7)	2669720 (100)	71.197782 (0.08)	85.1
ER0258	3154242 (1.6)	3082663 (97.7)	3082584 (100)	48.410613 (0.04)	68.9
ER0259	4490498 (2.2)	4399245 (98)	4399206 (100)	63.397433 (0)	83.9
ER0260	4120760 (2.1)	4036758 (98)	4036711 (100)	68.331358 (0.02)	82.7
ER0261	3351840 (1.7)	3278816 (97.8)	3278471 (100)	52.448055 (0.34)	79.6
ER0270	3505538 (1.8)	3415817 (97.4)	3415794 (100)	58.402948 (0.02)	84.2
ER0271	2341654 (1.2)	2290512 (97.8)	2290442 (100)	48.322114 (0.14)	79.1
ER0277	3683578 (1.8)	3587213 (97.4)	3587032 (100)	53.812907 (0.01)	76.2
ER0278	294000 (1.8)	287107 (97.7)	287105 (100)	60.50121 (0.03)	44.8
ER0279	3568011 (1.8)	3477383 (97.5)	3477352 (100)	67.082308 (0.01)	89.1
ER0280	3102475 (1.6)	3006098 (96.9)	3005348 (100)	58.997568 (0.06)	84.2
ER0281	3834430 (1.9)	3664627 (95.6)	3664459 (100)	56.366741 (0.01)	85.9
ER0308	3119640 (1.6)	3020552 (96.8)	3020402 (100)	61.949613 (0.03)	83.2
ER0323	2426809 (1.2)	2243376 (92.4)	2243269 (100)	66.586933 (0.01)	80
ER0324	1439000 (1.7)	1145592 (79.6)	1145475 (100)	70.761783 (0.03)	63.8

Table 2.1. (cont'd)

ER0325	3700987 (1.9)	3601773 (97.3)	3601692 (100)	56.972047 (0.03)	78.2
ER0326	3749755 (1.9)	3610077 (96.3)	3610009 (100)	77.306206 (0.08)	88.4
ER0327	185000 (1.6)	170720 (92.3)	170705 (100)	67.252863 (0.22)	24.4
ER0336	3511575 (1.8)	3418976 (97.4)	3418923 (100)	62.121956 (0.13)	72.5
ER0364	3999048 (2)	3904613 (97.6)	3904449 (100)	72.756066 (0.04)	95.9
ER0413	3103342 (1.6)	2770518 (89.3)	2727572 (98.4)	44.762937 (0.09)	81.6
ER0439	3178552 (1.6)	3061584 (96.3)	3061458 (100)	75.10139 (0.03)	84.8
ER0440	156000 (1.7)	145301 (93.1)	145299 (100)	85.724609 (0.01)	73.5
ER0446	3041907 (1.5)	2932061 (96.4)	2931818 (100)	68.021245 (0.02)	87.4
ER0465	3602899 (1.8)	3501480 (97.2)	3501370 (100)	60.392224 (0.03)	90.8
ER0466	2983258 (1.5)	2789012 (93.5)	2784280 (99.8)	66.697795 (2.39)	86.1
ER0467	3038648 (1.5)	2903454 (95.6)	2903139 (100)	61.470901 (0.05)	86.2
ER0468	2609413 (1.3)	2548530 (97.7)	2548505 (100)	72.31197 (0.06)	88.4
ER0469	3562669 (1.8)	3472148 (97.5)	3472071 (100)	71.4306 (0.02)	90.6
ER0470	3424225 (1.7)	3326276 (97.1)	3326229 (100)	73.126744 (0.01)	93.7
ER0490	2907025 (1.5)	2837080 (97.6)	2836967 (100)	70.320706 (0.01)	90
ER0499	3121529 (1.6)	3034411 (97.2)	3034089 (100)	68.255054 (0.01)	92.6
ER0501	3456873 (1.7)	3360751 (97.2)	3360518 (100)	66.725358 (0.01)	89.5
ER0503	2778570 (1.4)	2700540 (97.2)	2700286 (100)	67.444696 (0)	92.8
ER0516	2601615 (1.3)	2441463 (93.8)	2438927 (99.9)	67.301717 (18.65)	92.2
ER0541	3459582 (1.7)	3343595 (96.6)	3343582 (100)	58.343358 (0.35)	95.2
ER0561	3360491 (1.7)	3223756 (95.9)	3223649 (100)	65.327449 (0.09)	79.7
ER0612	3222010 (1.6)	3127004 (97.1)	3126934 (100)	64.859773 (0.37)	87.6
ER0626	3291513 (1.6)	3080888 (93.6)	3080431 (100)	63.887005 (0.04)	84.5
ER0627	3158973 (1.6)	2448164 (77.5)	2447774 (100)	58.972449 (0.13)	69.4
ER0634	2257316 (1.1)	2039657 (90.4)	2037943 (99.9)	65.067949 (0.24)	77.6
ER0664	2565189 (1.3)	2113453 (82.4)	2113142 (100)	57.720016 (0.14)	81.2
ER0671	3072584 (1.5)	2982100 (97.1)	2981986 (100)	69.241249 (0.02)	89.1
ER0690	3325404 (1.7)	3138520 (94.4)	3138246 (100)	65.935758 (0.01)	84.2

Table 2.1. (cont'd)

ER0691	2919300 (1.5)	2729918 (93.5)	2723876 (99.8)	78.275627 (0.01)	90
ER0692	3253856 (1.6)	3058749 (94)	3057723 (100)	82.487902 (0.22)	89.7
ER0698	3080854 (1.5)	2753514 (89.4)	2752848 (100)	53.110106 (0.05)	79.4
ER0699	4182037 (2.1)	4012817 (96)	4012757 (100)	57.69982 (0.08)	88.6
ER0709	3523923 (1.8)	3356211 (95.2)	3353466 (99.9)	61.937564 (0)	98.5
ER0739	3310926 (1.7)	3173864 (95.9)	3173802 (100)	58.05347 (0.09)	86.1
ER0741	3258579 (1.6)	3057965 (93.8)	3056360 (99.9)	45.415706 (0.38)	82.7
ER0763	3744482 (1.9)	3613376 (96.5)	3613220 (100)	70.974029 (0.01)	90.3
ER0780	3458005 (1.7)	3341789 (96.6)	3341574 (100)	67.861244 (0.04)	84.7
ER0781	3660688 (1.8)	3550589 (97)	3550541 (100)	65.110441 (0.01)	87
ER0797	3298042 (1.6)	3198885 (97)	3197924 (100)	57.594757 (0.12)	74.7
ER0886	3293716 (1.6)	3118915 (94.7)	3118785 (100)	49.57033 (1.77)	80.3
ER0887	3191976 (1.6)	3088536 (96.8)	3088522 (100)	61.784428 (0.04)	92.6
ER0944	3022773 (1.5)	2056838 (68)	2056703 (100)	51.88603 (0.12)	98.1
ER0947	3187266 (1.6)	3031943 (95.1)	3031458 (100)	73.457522 (0.26)	94.5
ER0958	2425696 (1.2)	2376242 (98)	2375942 (100)	70.169604 (0)	88.8
ER0959	2395695 (1.2)	2084851 (87)	2084501 (100)	60.83966 (0.01)	88
ER0961	3041780 (1.5)	2642845 (86.9)	2642585 (100)	65.16026 (0.04)	72.8
ER0964	2530594 (1.3)	2471368 (97.7)	2470375 (100)	50.508265 (0.01)	89.3
ER0974	3479256 (1.7)	3352624 (96.4)	3352448 (100)	76.237358 (0.05)	88.6
ER1005	3617399 (1.8)	3540560 (97.9)	3535046 (99.8)	59.87265 (0.34)	97.2
ER1010	590000 (1.6)	564765 (95.7)	564741 (100)	67.501858 (0)	80.6
ER1012	3211292 (1.6)	3086791 (96.1)	3085013 (99.9)	66.919134 (0.01)	88.2
ER1013	2400919 (1.2)	2233385 (93)	2233337 (100)	72.956246 (0.01)	82
ER1014	2827546 (1.4)	2724897 (96.4)	2724862 (100)	61.710791 (0.02)	82.4
ER1015	3146552 (1.6)	2917137 (92.7)	2916969 (100)	67.636418 (0.02)	81.3
ER1016	2437204 (1.2)	2317764 (95.1)	2317613 (100)	62.170517 (0.03)	78.5
ER1017	3097654 (1.5)	3014662 (97.3)	3014583 (100)	64.881034 (0.01)	83.1
ER0212	5106975 (2.6)	4838587 (94.7)	4838511 (100)	68.707943 (0.68)	89.2

Table 2.1. (cont'd)

ER0583	7427518 (3.7)	7248153 (97.6)	7247878 (100)	72.445071 (0)	97.7
ER0128	3488364 (1.7)	3405121 (97.6)	3405029 (100)	68.198068 (0.41)	82.2
ER0138	2809243 (1.4)	2529845 (90.1)	2529555 (100)	60.679996 (0)	89.8
ER0500	2826292 (1.4)	2764180 (97.8)	2758963 (99.8)	47.98185 (0.01)	89
ER1018	4840936 (2.4)	4722384 (97.6)	4722201 (100)	57.167467 (0.04)	84.2

Table 2.2. Characteristics of the 79 patients with enteric infections and 125 non-infected family members in the study

Characteristic	No. of cases‡	Percent (%) of cases	No. of non-infected ‡	Percent (%) of non-infected	*p*-value
Demographic data					
Sex					
Male	38	48.1	67	54.5	0.3765
Female	41	51.9	56	45.5	-
Age group (years)					
0-9	21	26.6	33	26.8	0.5497
10-18	11	13.9	17	13.8	0.6073
19-64	33	41.8	64	52	-
65+	14	17.7	9	7.3	0.0177
Race					
Caucasian	60	81.1	14	82.4	1.0
African American	10	13.5	3	17.6	0.5412
Other	4	5.4	0	0	-
Residence location					
Rural	40	51.3	55	47.4	0.5971
Urban	38	48.7	61	52.6	-
Residence (counties in Michigan)					
Calhoun	1	1.28	2	1.7	1
Clinton	4	5.13	3	2.6	0.6867
Eaton	5	6.41	13	11.2	0.3749
Ingham	16	20.51	15	12.9	0.628
Ionia	2	2.56	7	6	0.279
Kent	5	6.41	2	1.7	0.2404
Lenawee	1	1.28	3	2.6	0.6235
Livingston	3	3.85	5	4.3	1
Macomb	3	3.85	12	10.3	0.1233
Newaygo	0	0	4	3.4	0.1362
Oakland	8	10.26	20	17.2	0.2979
Ottawa	3	3.85	2	1.7	0.6486
Shiawassee	0	0	1	0.9	1
Washtenaw	11	14.1	7	6	0.3869
Wayne	16	20.51	20	17.2	-

Table 2.2. (cont'd)

Infection					
Camapylobacter	29	36.7	45	36	0.7483
Salmonella	35	44.3	57	45.6	0.7465
Shigella	10	12.7	17	13.6	0.7219
STEC	5	6.3	6	4.8	-
Epidemiological data					
Travel					
Domestic travel past 2 weeks					
Yes	16	21.3	34	30.6	0.1606
No	59	78.7	77	69.4	-
International travel past 2 weeks					
Yes*****	9	11.8	2	1.8	0.0080
No	67	88.2	109	98.2	-
Food consumption					
Turkey					
Yes	10	40	35	28.5	0.2526
No	15	60	88	71.5	-
Chicken					
Yes*	55	84.6	119	96.7	0.0060
No	10	15.4	4	3.3	-
Beef					
Yes	39	88.6	99	80.5	0.2552
No	5	11.4	24	19.5	-
Pork					
Yes*	33	82.5	60	48.8	0.0002
No	7	17.5	63	51.2	-
Deli meat					
Yes	25	51	69	56.1	0.5460
No	24	49	54	43.9	-
Raw fruits					
Yes	31	83.8	109	88.6	0.4356
No	6	16.2	14	11.4	-
Raw leafy greens					
Yes	36	67.9	87	70.7	0.7096
No	17	32.1	36	29.3	-

Table 2.2. (cont'd)

Raw vegetables					
Yes	21	63.6	95	77.2	0.6910
No	13	39.4	28	22.8	-
Raw eggs					
Yes	1	2.5	7	5.7	0.6807
No	39	97.5	116	94.3	-
Water at home					
Any well*	13	18.8	20	37.7	0.0021
Any municipal*	48	69.6	33	62.3	0.0228
Only bottled	8	11.6	0	0	-

The percentages based on the number for which information was available. Counts are mutually exclusive for each category. ‡ Total number varies due to the difference in missing data. * indicates significance difference ($p < 0.05$) between variables using *p*-value calculated by Chi-square test and Fisher's exact test for variables ≤ 5 in at least one cell. Mantel-Hanzel Chi-square was used to assess for trends.

Table 2.3. Clinical outcomes and animal contacts of the 79 patients with enteric infections included in the study

Characteristic	No. of cases‡	Percentage (%) of cases
Clinical Outcomes among cases only		
Case hospitalization		
Yes	29	37.2
No	49	62.8
Hospital Duration		
> 2 days	15	53.6
≤ 2 days	13	46.4
Abdominal pain		
Yes	12	15.6
No	65	84.4
Body ache		
Yes	22	28.6
No	55	71.4
Diarrhea		
Yes	73	94.8
No	4	5.2
Bloody diarrhea		
Yes	29	37.7
No	48	62.3
Chills		
Yes	25	32.5
No	52	67.5
Fatigue		
Yes	41	53.2
No	36	46.8
Headache		
Yes	18	23.4
No	59	76.6
Nausea		
Yes	38	49.4
No	39	50.6
Vomiting		
Yes	27	35.1
No	50	64.9
Fever		
Yes	47	69.1
No	21	30.9
Animal Contact		
Any animal		
Yes	46	63.9
No	26	36.1
Reptile		
Yes	5	6.9
No	67	93.1

Table 2.3. (cont'd)

Livestock		
Yes	10	13.9
No	62	86.1
Birds/poultry		
Yes	15	20.8
No	57	79.2
Domestic		
Yes	40	55.6
No	32	44.4
Others		
Yes	13	18.1
No	59	81.9

The percentages based on the number for which information was available. Counts are mutually exclusive for each category. ‡ Total number varies due to the difference in missing data.

Table 2.4. Characteristics of individuals with microbiome profiles belonging to one of the four Clusters defined through hierarchical clustering

Characteristic	Cluster 1‡ No. (%)	Cluster 2‡ No. (%)	Cluster 3‡ No. (%)	Cluster 4‡ No. (%)	*p*-value
Demographic data					
Case status					
Case	37 (74)	28 (96.6)	1 (2.3)	13 (16	< 0.0001
Control	13 (26)	1 (3.4)	43 (97.7)	68 (84)	-
Sex					
Male*	24 (48)	17 (58.6)	30 (69.8)	34 (42.5)	0.0264
Female	26 (52)	12 (41.4)	13 (30.2)	46 (57.5)	-
Age group (years)					
0-9	12 (24)	8 (27.6)	13 (30.2)	21 (26.3)	0.3958
10-18	7 (14)	4 (13.8)	6 (14)	11 (13.7)	0.6406
19-64	24 (48)	12 (41.4)	22 (51.1)	39 (38.7)	0.3460
65+	7 (14)	5 (17.2)	2 (4.7)	9 (11.3)	-
Race					
Caucasian	31 (81.6)	21 (80.8)	6 (100)	16 (76.2)	1
African American	5 (13.1)	4 (15.4)	0 (0)	4 (19.1)	1
Other	2 (5.3)	1 (3.8)	0 (0)	1 (4.7)	-
Residence type					
Rural*	26 (52)	13 (44.8)	26 (66.7)	30 (39.5)	0.0457
Urban	24 (48)	16 (55.2)	13 (33.3)	46 (60.5)	-
Infection					
Campylobacter	19 (38)	7 (24.1)	18 (40.9)	30 (37)	0.6008
Salmonella	20 (40)	14 (48.3)	21 (47.7)	37 (45.7)	0.4351
Shigella	6 (12)	7 (24.1)	3 (6.8)	11 (13.6)	0.4055
STEC	5 (10)	1 (3.4)	2 (4.5)	3 (3,7)	-
Epidemiological data					
Travel					
Domestic travel					
Yes*	10 (20.8)	9 (33.3)	19 (45.2)	12 (17.4)	0.0080
No	38 (79.2)	18 (66.7)	23 (54.8)	57 (82.6)	-
International travel					
Yes	4 (8.2)	3 (11.1)	1 (2.4)	3 (4.3)	0.3372
No	45 (91.8)	24 (88.9)	41 (97.6)	66 (95.7)	-

The percentages based on the number for which information was available. Counts are mutually exclusive for each category. ‡ Total number varies due to the difference in missing data. * indicates significance difference ($p < 0.05$) between variables using *p*-value calculated by Chi-square test and Fisher's exact test for variables ≤ 5 in at least one cell. Mantel-Hanzel Chi-square was used to assess for trends.

Table 2.5. Univariate analysis to identify disease associations for Cluster 1 in 79 patients with enteric infections

Outcome	Totals*	No (%) Cluster 1	OR (95% CI)†	*p*-value‡
Sex				
Male	38	15 (39.5)	1.0	-
Female	41	22 (53.7)	0.6 (0.2 - 1.4)	0.2068
Age group (years)				
0-9	21	11 (47.6)	0.9 (0.3 - 2.7)	0.8764
10-18	11	5 (45.4)	1 (0.3 - 3.9)	1
19-64	33	15 (45.4)	1.0	-
65+	14	7 (50)	0.8 (0.2 - 2.9)	0.7752
Race				
Caucasian	60	28 (46.7)	1.1 (0.1 – 16)	0.9
African American	10	5 (50)	1 (0.1 – 19)	1
Other	4	2 (50)	1.0	-
Residence Type				
Urban	38	18 (47.4)	1 (0.4 - 2.4)	0.9907
Rural	40	19 (47.5)	1.0	-
Infection				
Campylobacter	29	15 (46.8)	1.4 (0.1 - 19)	0.7320
Salmonella	35	17 (48.6)	1.6 (0.2 - 21)	0.6326
Shigella	10	2 (20)	5 (0.3 - 111)	0.1213
STEC	5	3 (60)	1.0	-
Hospitalized				
Yes	29	13 (44.8)	0.8 (0.3 - 2.1)	0.7227
No	49	24 (49)	1.0	-
Abdominal pain				
Yes	65	30 (46.7)	0.6 (0.1 - 2.5)	0.5360
No	12	7 (58.3)	1.0	-
Body ache				
Yes	22	16 (72.7)	4.3 (1.5 - 12.8)	0.0061
No	55	21 (38.2)	1.0	-
Diarrhea				
Yes	73	36 (49.3)	2.9 (0.2 - 157)	0.6161
No	4	1 (25)	1.0	-
Bloody diarrhea				
Yes	29	10 (34.5)	0.4 (0.2 - 1.1) 1.0	0.0639
No	48	27 (56.2)		-
Chills				
Yes	25	12 (48)	1 (0.4 - 2.6)	0.9949
No	52	25 (48.1)	1.0	-

Table 2.5. (cont'd)

Fatigue				
Yes	41	18 (43.9)	0.7 (0.3 - 1.7)	0.4367
No	36	19 (52.8)	1.0	-
Headache				
Yes	18	9 (50)	1.1 (0.4 - 3.2)	0.8501
No	59	28 (47.5)	1.0	-
Nausea				
Yes	38	19 (50)	1.2 (0.5 - 2.9)	0.7356
No	39	18 (46.2)	1.0	-
Vomiting				
Yes	27	12 (44.4)	0.8 (0.3 - 2)	0.6415
No	50	25 (50)	1.0	-
Fever				
Yes	47	20 (42.6)	0.6 (0.2 - 1.6)	0.2654
No	21	12 (57.1)	1.0	-

* Depending on the variable examined, the number does not add up to the total (n=79) because of missing data. † 95% confidence interval (CI) for odds ratio (OR). ‡ *p*-value calculated by Chi-square test and Fisher's exact test was used for variables $\leq$5 in at least one cell. Mantel-Hanzel Chi-square was used to assess for trends.

Table 2.6. Univariate and multivariate analysis to identify disease associations for Cluster 2 in 79 patients

Characteristic	Total*	No (%) Cluster 2	OR (95% CI)†	*p*-value‡
Sex				
Male	38	16 (42.1)	1.0	-
Female	41	12 (29.3)	1.8 (0.7 - 4.5)	0.2333
Age group (years)				
0-9	21	8 (28.6)	0.8 (0.2 - 3)	0.7753
10-18	11	4 (14.3)	0.9 (0.2 - 5)	1
19-64	33	11 (39.3)	1.0	-
65+	14	5 (17.9)	0.9 (0.2 - 4.3)	1
Race				
Caucasian	60	21 (35)	0.5 (0.01 - 8.3)	1
African American	10	4 (40)	0.8 (0.1 - 9.8)	1
Other	4	1 (25)	1.0	-
Residence Type				
Urban	38	16 (42.1)	1.7 (0.7 - 4.3)	0.2653
Rural	40	12 (30)	1.0	-
Infection				
Campylobacter	29	7 (24.1)	0.8 (0.01 - 10)	1
Salmonella	35	13 (37.1)	0.4 (0.01 - 5)	0.6404
Shigella	10	7 (70)	0.1(0.001 - 2)	0.1189
STEC	5	1 (20)	1.0	-
Hospitalized				
Yes	29	11 (37.9)	1.2 (0.4 - 3)	0.7733
No	49	17 (34.7)	1.0	-
Abdominal pain				
Yes	65	23 (35.4)	1.6 (0.4 - 10.3)	0.7410
No	12	3 (25)	1.0	-
Body ache				
Yes	22	5 (22.7)	0.5 (0.1 - 1.6)	0.2867
No	55	21 (38.2)	1.0	-
Diarrhea				
Yes	73	26 (35.6)	Un (0.3 – Un)	0.2937
No	4	0 (0)	1.0	-
Bloody diarrhea				
Yes	29	15 (51.7)	3.6 (1.3 - 9.7)	0.0096
No	48	11 (22.9)	1.0	-
Chills				
Yes	25	8 (32)	0.9 (0.3 - 2.5)	0.8202
No	52	18 (34.6)	1.0	-

Table 2.6. (cont'd)

Fatigue				
Yes	41	16 (39)	1.7 (0.6 - 4.4)	0.2978
No	36	10 (27.8)	1.0	-
Headache				
Yes	18	9 (50)	2.5 (0.8 - 7.3)	0.0962
No	59	17 (28.8)	1.0	-
Nausea				
Yes	38	13 (34.2)	1 (0.4 - 2.7)	0.9351
No	39	13 (33.3)	1.0	-
Vomiting				
Yes	27	13 (48.1)	2.6 (1 - 7.1)	0.0499
No	50	13 (26)	1.0	-
Fever				
Yes	47	19 (40.4)	2.1 (0.6 - 8.8)	0.2728
No	21	5 (23.8)	1.0	-

* Depending on the variable examined, the number does not add up to the total (n=79) because of missing data. † 95% confidence interval (CI) for odds ratio (OR). ‡ *p*-value calculated by Chi-square test and Fisher's exact test was used for variables $\leq$5 in at least one cell. Mantel-Hanzel Chi-square was used to assess for trends.

Table 2.7. Differentially abundant taxa determined by ANCOM for each case cluster

Organism (Genus)	Taxonomy (Order; Family)	Cluster 1	Cluster 2
Viruses			
P22virus	Caudovirales; Podoviridae		Present
P2virus	Caudovirales; Myoviridae		Present
Nona33virus	Caudovirales; Podoviridae		Present
Lambdavirus	Caudovirales; Siphoviridae		Present
Kp15virus	Caudovirales; Myoviridae		Present
Hk97virus	Caudovirales; Siphoviridae		Present
P1virus	Caudovirales; Myoviridae		Present
T7virus	Caudovirales; Podoviridae		Present
Sk1virus	Caudovirales; Siphoviridae		Present
L5virus	Caudovirales; Siphoviridae		Present
Felixo1virus	Caudovirales; Myoviridae		Present
C5virus	Caudovirales; Siphoviridae		Present
Epsilon15virus	Caudovirales; Podoviridae		Present
Jerseyvirus	Caudovirales; Siphoviridae		Present
Pepy6virus	Caudovirales; Siphoviridae		Present
T5virus	Caudovirales; Siphoviridae		Present
Sfi11virus	Caudovirales; Siphoviridae		Present
Pis4avirus	Caudovirales; Siphoviridae		Present
Muvirus	Caudovirales; Myoviridae		Present
Sfi21dt1virus	Caudovirales; Siphoviridae		Present
K1gvirus	Caudovirales; Siphoviridae		Present
Cytomegalovirus	Herpesvirales; Herpesviridae		Present
Tl2011virus	Caudovirales; Podoviridae		Present
Rb69virus	Caudovirales; Myoviridae		Present
Jd18virus	Caudovirales; Myoviridae		Present
S16virus	Caudovirales; Myoviridae		Present
Orthopoxvirus	Viruses; Poxviridae	Present	Present
Bacteria			
Salmonella	Enterobacterales; Enterobacteriaceae	Present	Present
Alistipes	Bacteroidales; Rikenellaceae	Present	Present
Escherichia	Enterobacterales; Enterobacteriaceae	Present	Present
Shigella	Enterobacterales; Enterobacteriaceae	Present	Present
Klebsiella	Enterobacterales; Enterobacteriaceae	Present	Present
Enterobacter	Enterobacterales; Enterobacteriaceae	Present	Present
Citrobacter	Enterobacterales; Enterobacteriaceae	Present	Present
Haemophilus	Enterobacterales; Enterobacteriaceae	Present	Present
Oscillibacter	Clostridiales; Oscillospiraceae	Present	Present
Serratia	Enterobacterales; Yersiniaceae	Present	Present
Atlantibacter	Enterobacterales; Enterobacteriaceae	Present	Present
Raoultella	Enterobacterales; Enterobacteriaceae	Present	Present
Kluyvera	Enterobacterales; Enterobacteriaceae	Present	Present
Proteus	Enterobacterales; Morganellaceae	Present	Present

Table 2.7. (cont'd)

Hafnia	Enterobacterales; Hafniaceae	Present	Present
Neglecta	Clostridiales; Ruminococcaceae	Present	Present
Morganella	Enterobacterales; Morganellaceae	Present	Present
Bacteroides	Bacteroidales; Bacteroidaceae		Present
Roseburia	Clostridiales; Lachnospiraceae		Present
Clostridioides	Clostridiales; Peptostreptococcaceae		Present
Ruminococcus	Clostridiales; Ruminococcaceae		Present
Butyricicoccus	Clostridiales; Clostridiaceae		Present
Chlamydia	Chlamydiales; Chlamydiaceae		Present
Porphyromonas	Bacteroidales; Porphyromonadaceae		Present
Eubacterium	Clostridiales; Eubacteriaceae		Present
Lactococcus	Lactobacillales; Streptococcaceae		Present
Streptococcus	Lactobacillales; Streptococcaceae		Present
Flavonifractor	Clostridiales; Ruminococcaceae		Present
Holdemania	Erysipelotrichales; Erysipelotrichaceae		Present
Subdoligranulum	Clostridiales; Ruminococcaceae		Present
Azospirillum	Rhodospirillales; Rhodospirillaceae		Present
Tannerella	Bacteroidales; Tannerellaceae		Present
Anaerotruncus	Clostridiales; Ruminococcaceae		Present
Agathobaculum	Clostridiales; Ruminococcaceae		Present
Dysgonomonas	Bacteroidales; Dysgonamonadaceae		Present
Fusicatenibacter	Clostridiales; Lachnospiraceae		Present
Acinetobacter	Pseudomonadales; Moraxellaceae		Present
Pseudomonas	Pseudomonadales; Pseudomonadaceae		Present
Pseudoflavonifractor	Clostridiales; Ruminococcaceae		Present
Staphylococcus	Bacillales; Staphylococcaceae		Present
Bacillus	Bacillales; Bacillaceae		Present
Enterococcus	Lactobacillales; Enterococcaceae		Present
Lactobacillus	Lactobacillales; Lactobacillaceae		Present
Alloprevotella	Bacteroidales; Prevotellaceae		Present
Paenibacillus	Bacillales; Paenibacillaceae		Present
Intestinibacillus	Clostridiales; Eubacteriaceae		Present
Intestinimonas	Clostridiales; unclassified Clostridiales		Present
Ruthenibacterium	Clostridiales; Ruminococcaceae		Present
Gemmiger	Clostridiales; Ruminococcaceae		Present
Anaeromassilibacillus	Clostridiales; Ruminococcaceae		Present
Angelakisella	Clostridiales; Ruminococcaceae		Present
Lawsonibacter	Clostridiales; unclassified Clostridiales		Present
Lachnotalea	Clostridiales; Lachnospiraceae		Present
Peptostreptococcus	Clostridiales; Peptostreptococcaceae		Present
Acetobacter	Rhodospirillales; Acetobacteraceae		Present
Acidovorax	Burkholderiales; Comamonadaceae		Present
Colibacter	Veillonellales; Veillonellaceae		Present
Tissierella	Tissierellales; Tissierellaceae	Present	Present

Table 2.7. (cont'd)

Victivallis	Victivallales; Victivallaceae	Present	Present
Lachnoclostridium	Clostridiales; Lachnospiraceae	Present	
Akkermansia	Verrucomicrobiales; Akkermansiaceae	Present	
Veillonella	Veillonellales; Veillonellaceae	Present	
Odoribacter	Bacteroidales; Odoribacteraceae		
Christensenella	Clostridiales; Christensenellaceae		
Asaccharobacter	Eggerthellales; Eggerthellaceae	Present	

Table 2.8. Univariate and multivariate analysis for Cluster 2 status in 79 patients with enteric infections and 125 non-infected family members (controls) included in the study

Characteristic	Total*	No (%) Cluster 2	OR (95% CI) †	*p*-value‡
Viruses above study average				
P22virus	15	7 (46.7)	6.6 (2.2 - 20.1)	0.0001
Yes	189	22 (11.6)	1.0	-
No				
P2virus	17	9 (52.9)	9.4 (3.3 - 27.1)	< 0.0001
Yes	187	20 (10.7)	1.0	-
No				
Nona33virus	21	13 (61.9)	17 (6.1 - 47)	< 0.0001
Yes	183	16 (8.7)	1.0	-
No				
Lambda	19	9 (47.4)	7.4 (2.7 - 20.4)	0.0000
Yes	185	20 (10.8)	1.0	-
No				
Orthopoxvirus	38	20 (52.6)	19.4 (7.7 - 48.9)	0.0000
Yes	166	9 (5.4)	1.0	-
No				
Kp15virus	2	1 (50)	6.1 (0.1 - 488.4)	0.2647
Yes	202	28 (13.9)	1.0	-
No				
Hk97virus	25	11 (44)	7 (2.8 - 17.8)	0.0000
Yes	179	18 (10.1)	1.0	-
No				
P1virus	12	6 (50)	7.3 (2.2 - 24.7)	0.0002
Yes	192	23 (12)	1.0	-
No				
T7virus	14	0 (0)	0 (0 - 1.8)	0.2270
Yes	190	29 (15.3)	1.0	-
No				
Sk1virus	11	1 (9.1)	0.6 (0 - 4.5)	1
Yes	193	28 (14.5)	1.0	-
No				
L5virus	34	0 (0)	0 (0 - 0.6)	0.0056
Yes	170	29 (17.1)	1.0	-
No				
Felixo1virus	3	1 (33.3)	3.1 (0.1 - 60.8)	0.3702
Yes	201	28 (13.9)	1.0	-
No				
C5virus	10	0 (0)	0 (0 - 2.7)	0.3630
Yes	194	29 (14.9)	1.0	-
No				

Table 2.8. (cont'd)

Epsilon15virus				
Yes	10	4 (40)	4.5 (0.9 - 20.3)	0.0378
No	194	25 (12.9)	1.0	-
Jerseyvirus				
Yes	6	1 (16.7)	1.2 (0 - 11.4)	1
No	198	28 (14.1)	1.0	-
Pepy6virus				
Yes	17	0 (0)	0 (0 - 1.4)	0.1388
No	187	29 (15.5)	1.0	-
T5virus				
Yes	2	2 (100)	Inf (1.2 - Inf)	0.0196
No	202	27 (13.4)	1.0	-
Sfi11virus				
Yes	19	0 (0)	0 (0 - 1.2)	0.0817
No	185	29 (15.7)	1.0	-
Pis4avirus				
Yes	4	2 (50)	6.3 (0.4 - 90.6)	0.0976
No	200	27 (13.5)	1.0	-
Muvirus				
Yes	7	2 (28.6)	2.5 (0.2 - 16.3)	0.2605
No	197	27 (13.7)	1.0	-
Sfi21dt1virus				
Yes	19	0 (0)	0 (0 - 1.2)	0.0817
No	185	29 (15.7)	1.0	-
K1gvirus				
Yes	3	1 (33.3)	3.1 (0.1 - 60.8)	0.3702
No	201	28 (13.9)	1.0	-
Cytomegalovirus				
Yes	20	6 (30)	3 (1 - 8.6)	0.0333
No	184	23 (12.5)	1.0	-
Tl2011virus				
Yes	4	2 (50)	6.3 (0.4 - 90.6)	0.0976
No	200	27 (13.5)	1.0	-
Rb69virus				
Yes	4	0 (0)	0 (0 - 9.3)	1
No	200	29 (14.5)	1.0	-
Jd18virus				
Yes	1	0 (0)	0 (0 - 234.5)	1
No	203	29 (14.3)	1.0	-
S16virus				
Yes	5	0 (0)	0 (0 - 6.7)	1
No	199	29 (14.6)	1.0	-

Table 2.8. (cont'd)

Bacteria above study average				
Bacteroides				
Yes	112	12 (10.7)	0.5 (0.2 - 1.2)	0.1140
No	92	17 (18.5)	1.0	-
Salmonella				
Yes	19	10 (52.6)	9.7 (3.5 - 26.9)	0.0000
No	185	19 (10.3)	1.0	-
Alistipes				
Yes	76	0 (0)	0 (0 - 0.2)	0.0000
No	128	29 (22.7)	1.0	-
Escherichia				
Yes	42	19 (45.2)	12.6 (5.2 - 30.3)	0
No	162	10 (6.2)	1.0	-
Roseburia				
Yes	47	0 (0)	0 (0 - 0.4)	0.0005
No	157	29 (18.5)	1.0	-
Shigella				
Yes	38	20 (52.6)	19.4 (7.7 - 48.9)	0
No	166	9 (5.4)	1.0	-
Clostridioides				
Yes	85	6 (7.1)	0.3 (0.1 - 0.8)	0.0133
No	119	23 (19.3)	1.0	-
Klebsiella				
Yes	27	8 (29.6)	3.1 (1.2 - 8)	0.0138
No	177	21 (11.9)	1.0	-
Ruminococcus				
Yes	63	0 (0)	0 (0 - 0.3)	0.0000
No	141	29 (20.6)	1.0	-
Enterobacter				
Yes	33	17 (51.5)	14.1 (5.7 - 34.6)	0
No	171	12 (7)	1.0	-
Butyricicoccus				
Yes	53	0 (0)	0 (0 - 0.3)	0.0001
No	151	29 (19.2)	1.0	-
Citrobacter				
Yes	21	10 (47.6)	7.8 (2.9 - 20.9)	0.0000
No	183	19 (10.4)	1.0	-
Chlamydia				
Yes	30	5 (16.7)	1.2 (0.3 - 3.8)	0.7765
No	174	24 (13.8)	1.0	-
Porphyromonas				
Yes	31	3 (9.7)	0.6 (0.1 - 2.2)	0.5812
No	173	26 (15)	1.0	-

Table 2.8. (cont'd)

Eubacterium				
Yes	51	0 (0)	0 (0 - 0.3)	0.0001
No	153	29 (19)	1.0	-
Lactococcus				
Yes	75	4 (5.3)	0.2 (0.1 - 0.7)	0.0061
No	129	25 (19.4)	1.0	-
Streptococcus				
Yes	21	11 (52.4)	10.1 (3.8 - 27)	0.0000
No	183	18 (9.8)	1.0	-
Flavonifractor				
Yes	64	0 (0)	0 (0 - 0.2)	0.0000
No	140	29 (20.7)	1.0	-
Haemophilus				
Yes	22	11 (50)	9.1 (3.5 - 24)	0.0000
No	182	18 (9.9)	1.0	-
Holdemania				
Yes	68	0 (0)	0 (0 - 0.2)	0.0000
No	136	29 (21.3)	1.0	-
Subdoligranulum				
Yes	68	0 (0)	0 (0 - 0.2)	0.0000
No	136	29 (21.3)	1.0	-
Azospirillum				
Yes	12	0 (0)	0 (0 - 2.2)	0.2226
No	192	29 (15.1)	1.0	-
Tannerella				
Yes	49	3 (6.1)	0.3 (0.1 - 1.1)	0.0976
No	155	26 (16.8)	1.0	-
Anaerotruncus				
Yes	65	0 (0)	0 (0 - 0.2)	0.0000
No	139	29 (20.9)	1.0	-
Agathobaculum				
Yes	65	0 (0)	0 (0 - 0.2)	0.0000
No	139	29 (20.9)	1.0	-
Dysgonomonas				
Yes	3	0 (0)	0 (0 - 14.9)	1
No	201	29 (14.4)	1.0	-
Fusicatenibacter				
Yes	51	0 (0)	0 (0 - 0.3)	0.0001
No	153	29 (19)	1.0	-
Acinetobacter				
Yes	22	14 (63.6)	19.5 (7 - 53.9)	0
No	182	15 (8.2)	1.0	-

Table 2.8. (cont'd)

Pseudomonas				
Yes	10	6 (60)	10.9 (2.4 - 56.7)	0.0007
No	194	23 (11.9)	1.0	-
Pseudoflavonifractor				
Yes	56	0 (0)	0 (0 - 0.3)	0.0000
No	148	29 (19.6)	1.0	-
Staphylococcus				
Yes	24	13 (54.2)	12.1 (4.7 - 31.4)	0.0000
No	180	16 (8.9)	1.0	-
Oscillibacter				
Yes	58	0 (0)	0 (0 - 0.3)	0.0000
No	146	29 (19.9)	1.0	-
Serratia				
Yes	22	15 (68.2)	25.7 (9 - 73.5)	0
No	182	14 (7.7)	1.0	-
Bacillus				
Yes	57	13 (22.8)	2.4 (1.1 - 5.4)	0.0286
No	147	16 (10.9)	1.0	-
Enterococcus				
Yes	20	14 (70)	26.3 (8.8 - 78.4)	0
No	184	15 (8.2)	1.0	-
Lactobacillus				
Yes	45	6 (13.3)	0.9 (0.3 - 2.4)	0.8477
No	159	23 (14.5)	1.0	-
Alloprevotella				
Yes	25	2 (8)	0.5 (0.1 - 2.2)	0.5412
No	179	27 (15.1)	1.0	-
Paenibacillus				
Yes	39	11 (28.2)	3.2 (1.4 - 7.5)	0.0054
No	165	18 (10.9)	1.0	-
Intestinibacillus				
Yes	53	0 (0)	0 (0 - 0.3)	0.0001
No	151	29 (19.2)	1.0	-
Intestinimonas				
Yes	63	0 (0)	0 (0 - 0.3)	0.0000
No	141	29 (20.6)	1.0	-
Ruthenibacterium				
Yes	43	0 (0)	0 (0 - 0.4)	0.0009
No	161	29 (18)	1.0	-
Atlantibacter				
Yes	4	3 (75)	20 (1.5 - 1059.6)	0.0094
No	200	26 (13)	1.0	-

Table 2.8. (cont'd)

Raoultella				
Yes	11	5 (45.5)	5.8 (1.3 - 24.8)	0.0103
No	193	24 (12.4)	1.0	-
Gemmiger				
Yes	58	0 (0)	0 (0 - 0.3)	0.0000
No	146	29 (19.9)	1.0	-
Anaeromassilibacillus				
Yes	56	0 (0)	0 (0 - 0.3)	0.0000
No	148	29 (19.6)	1.0	-
Kluyvera				
Yes	4	0 (0)	0 (0 - 9.3)	1
No	200	29 (14.5)	1.0	-
Angelakisella				
Yes	46	0 (0)	0 (0 - 0.4)	0.0004
No	158	29 (18.4)	1.0	-
Lawsonibacter				
Yes	46	0 (0)	0 (0 - 0.4)	0.0004
No	158	29 (18.4)	1.0	-
Lachnotalea				
Yes	56	0 (0)	0 (0 - 0.3)	0.0000
No	148	29 (19.6)	1.0	-
Peptostreptococcus				
Yes	10	7 (70)	17.8 (3.7 - 114.4)	0.0000
No	194	22 (11.3)	1.0	-
Proteus				
Yes	2	1 (50)	6.1 (0.1 - 488.4)	0.2647
No	202	28 (13.9)	1.0	-
Acetobacter				
Yes	15	0 (0)	0 (0 - 1.6)	0.1355
No	189	29 (15.3)	1.0	-
Hafnia				
Yes	4	1 (25)	2 (0 - 26.4)	0.4611
No	200	28 (14)	1.0	-
Neglecta				
Yes	41	0 (0)	0 (0 - 0.5)	0.0017
No	163	29 (17.8)	1.0	-
Morganella				
Yes	3	1 (33.3)	3.1 (0.1 - 60.8)	0.3702
No	201	28 (13.9)	1.0	-
Acidovorax				
Yes	3	1 (33.3)	3.1 (0.1 - 60.8)	0.3702
No	201	28 (13.9)	1.0	-

Table 2.8. (cont'd)

Colibacter	2	0 (0)	0 (0 - 32.5)	1
Yes	202	29 (14.4)	1.0	-
No				
Tissierella	2	0 (0)	0 (0 - 32.5)	1
Yes	202	29 (14.4)	1.0	-
No				
Victivallis	12	0 (0)	0 (0 - 2.2)	0.2226
Yes	192	29 (15.1)	1.0	-
No				

	Multivariate Analysis		
Logistic Regression	**OR**	**95 CI €**	***p* value‡**
Model 1			
Salmonella:			
Above study average: Yes	9.7	3.5 - 26.9	< 0.0001
Model 2			
Orthopoxvirus:			
Above study average: Yes	33.8	10.4 - 110.2	< 0.0001
Salmonella:			
Above study average: Yes	23.5	5.8 - 95.7	< 0.0001
Model 3			
Acinetobacter:			
Above study average: Yes	13.3	3.2 - 55.3	0.0005
Orthopoxvirus:			
Above study average: Yes	31.1	8.5 - 113.2	< 0.0001
Salmonella:			
Above study average: Yes	19.0	3.9 - 91.5	0.0003
Model 4			
Acinetobacter:			
Above study average: Yes	12.3	2.7 - 56.4	0.001
Orthopoxvirus:			
Above study average: Yes	14.4	3.5 - 58.5	0.0002
Salmonella:			
Above study average: Yes	26.2	5.0 - 136.6	0.0001
Serratia:			
Above study average: Yes	13.1	3.0 - 57.5	0.0008
Model 5			
Acinetobacter:			
Above study average: Yes	13.7	2.7 - 69.0	0.002
Orthopoxvirus:			
Above study average: Yes	16.7	3.4 - 81	0.0006
Salmonella:			
Above study average: Yes	28.2	5.2 - 154.0	0.0001

Table 2.8. (cont'd)

Serratia:			
Above study average: Yes	14.5	3.1 - 68.9	0.0009
Nona33virus:			
Above study average: Yes	0.7	0.1 - 4.0	0.7
Model 6			
Enterococcus:			
Above study average: Yes	2.5	0.3 - 18.3	0.4
Acinetobacter:			
Above study average: Yes	7.2	1.0 - 51.2	0.1
Orthopoxvirus:			
Above study average: Yes	13.9	3.4 - 57.3	0.0004
Salmonella:			
Above study average: Yes	23.2	4.2 - 127.4	0.0004
Serratia:			
Above study average: Yes	10.8	2.3 - 49.8	0.003
Model 7			
Enterococcus:			
Above study average: Yes	9.1	1.9 - 43.0	0.006
Orthopoxvirus:			
Above study average: Yes	16.3	4.1 - 65.0	0.0001
Salmonella:			
Above study average: Yes	23.6	4.7 - 118.6	0.0002
Serratia:			
Above study average: Yes	8.5	2.0 - 35.4	0.004
Model 8			
Acinetobacter:			
Above study average: Yes	12.3	2.7 - 56.4	0.001
Orthopoxvirus:			
Above study average: Yes	14.4	3.5 - 58.5	0.0003
Salmonella:			
Above study average: Yes	26.2	5.0 - 136.6	0.0001
Serratia:			
Above study average: Yes	13.1	3.0 - 57.5	0.0008

	Model Performance		
	Accuracy	**Accuracy (95 CI)**	**AUC**
Final Model (Model 8)	0.902	(0.7859, 0.9674)	0.9757

* The number of isolates may not add up to the total (n=204) due to missing data.

† 95 confidence interval (CI) for the odds ratio (OR)

Table 2.8. (cont'd)

‡ *p*-value was calculated by Chi-square and Fisher's exact test was used for variables $\leq$5 in at least one of the cells.

£ Logistic regression was performed via forward, backward selection while controlling for variables that yielded strong ($p \leq 0.20$) associations with the outcome as Cluster 2 in the univariate analysis. Hosmer and Lemeshow Goodness-of-Fit test was used to assess each model. All variables were tested for collinearity. € Wald 95 confidence intervals (CI).

Figure 2.1. Assessment of differences in microbiome profiles generated from samples sequenced using two different platforms

The principle component analysis (PCA) shows clustering of healthy individuals (controls; blue circles) and patients with enteric infections (cases; red circles) by infection type using the: A) Hiseq 2500v1; B) Hiseq 2500v2. C) All samples sequenced using both platforms were merged, and D) samples were stratified by sequencer. Elipses are CI (95%).

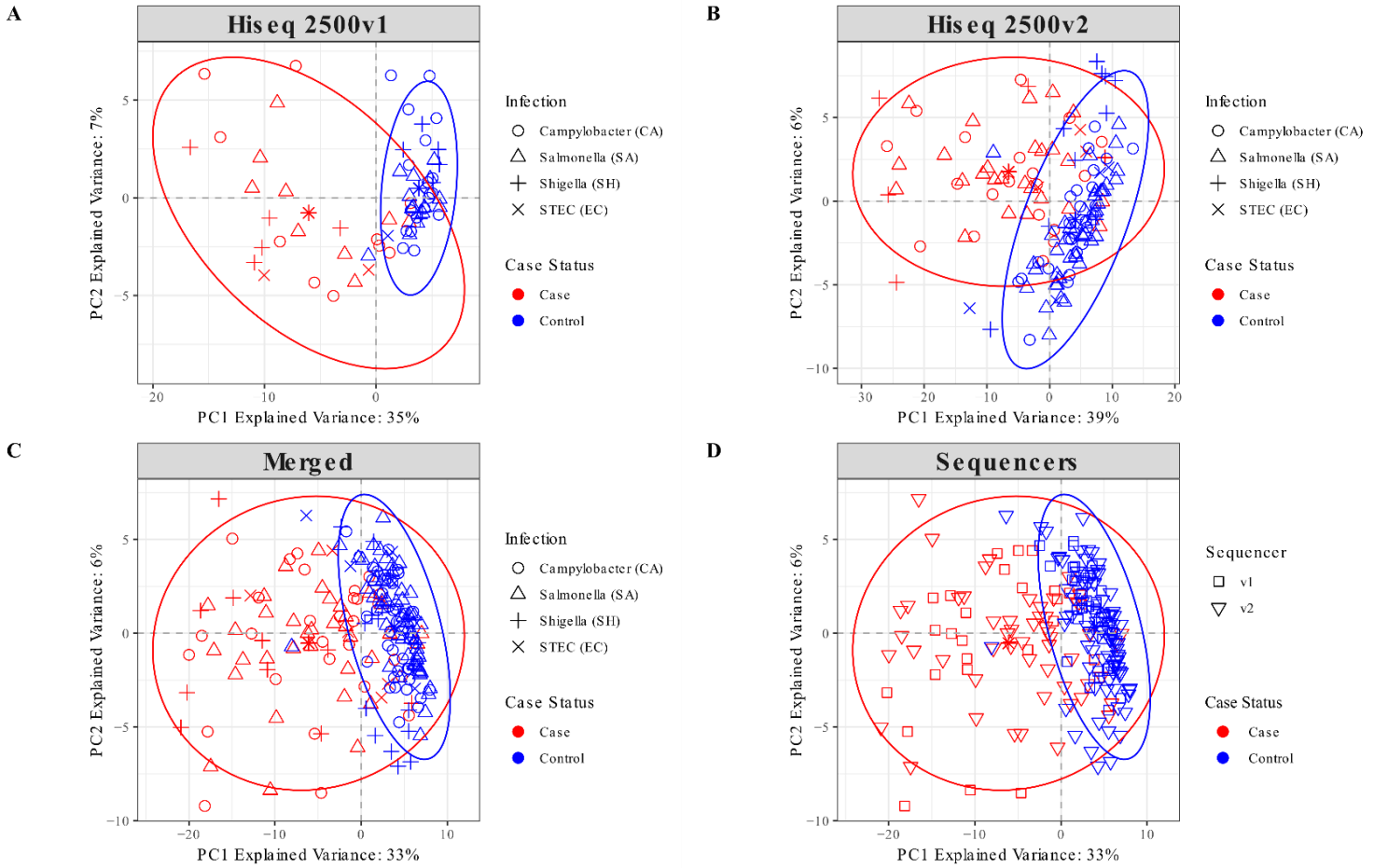


Figure 2.2. Power analysis demonstrating the sample size needed to detect differences between sample groups (cases versus controls) Power curves were created based on the original Cohen power equations using conventional parameters. The curves show the relationship between the effect size (differences in means over pooled standard deviations) and the sample size needed to detect that effect size. The black circle represents the sample size (n=204) used in our study, which has a ≤ 0.18 effect size and falls on the 0.8 power curve (blue line). Additional power curves at 0.5, 0.6, 0.7, and 0.9 were generated to yield different sample size and effect size estimates.

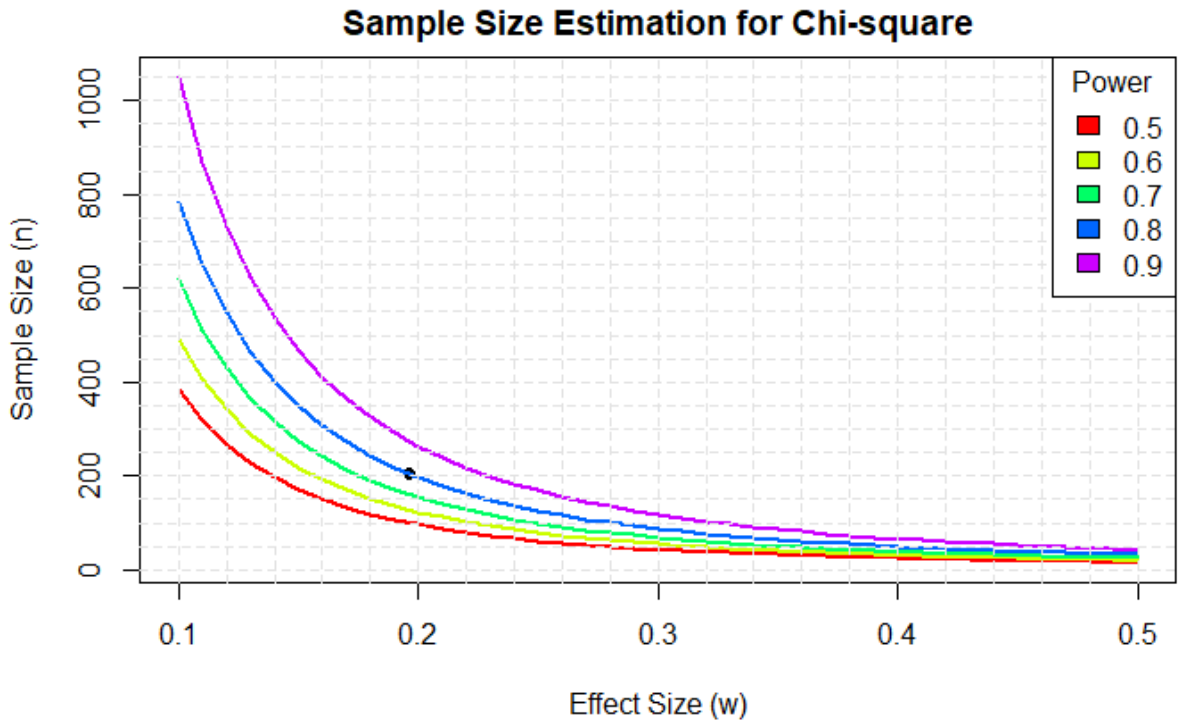


Figure 2.3. The percentage of bacterial and viral reads annotated at four taxonomical levels The number of reads annotated was compared to the number of quality-controlled reads for annotation and visualized based on the taxonomical level for all 204 samples. Annotated reads represent bacterial and viral sequences combined. For each taxonomical level, the line in the box represents the median, while the interquartile range (25%-75%) is the box surrounding the median. The whiskers indicate the variability outside the upper and lower quartiles, extending from 5%-95% of the samples. Outliers are represented as circles.

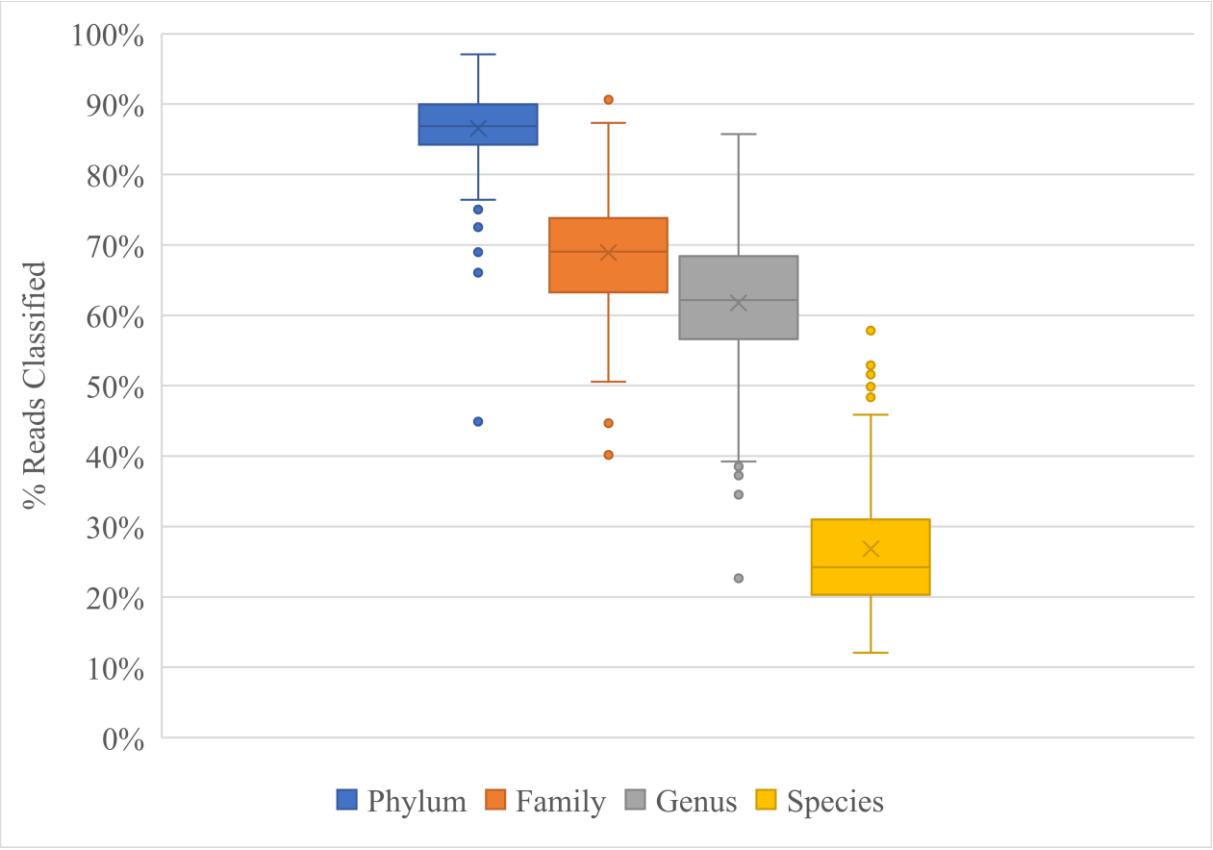


Figure 2.4. Principal Component Analysis (PCA) for 79 cases and 125 controls by infection type A) Order; B) Family; C) Genus; and D) Species. Ellipses indicate the 95% confidence intervals.

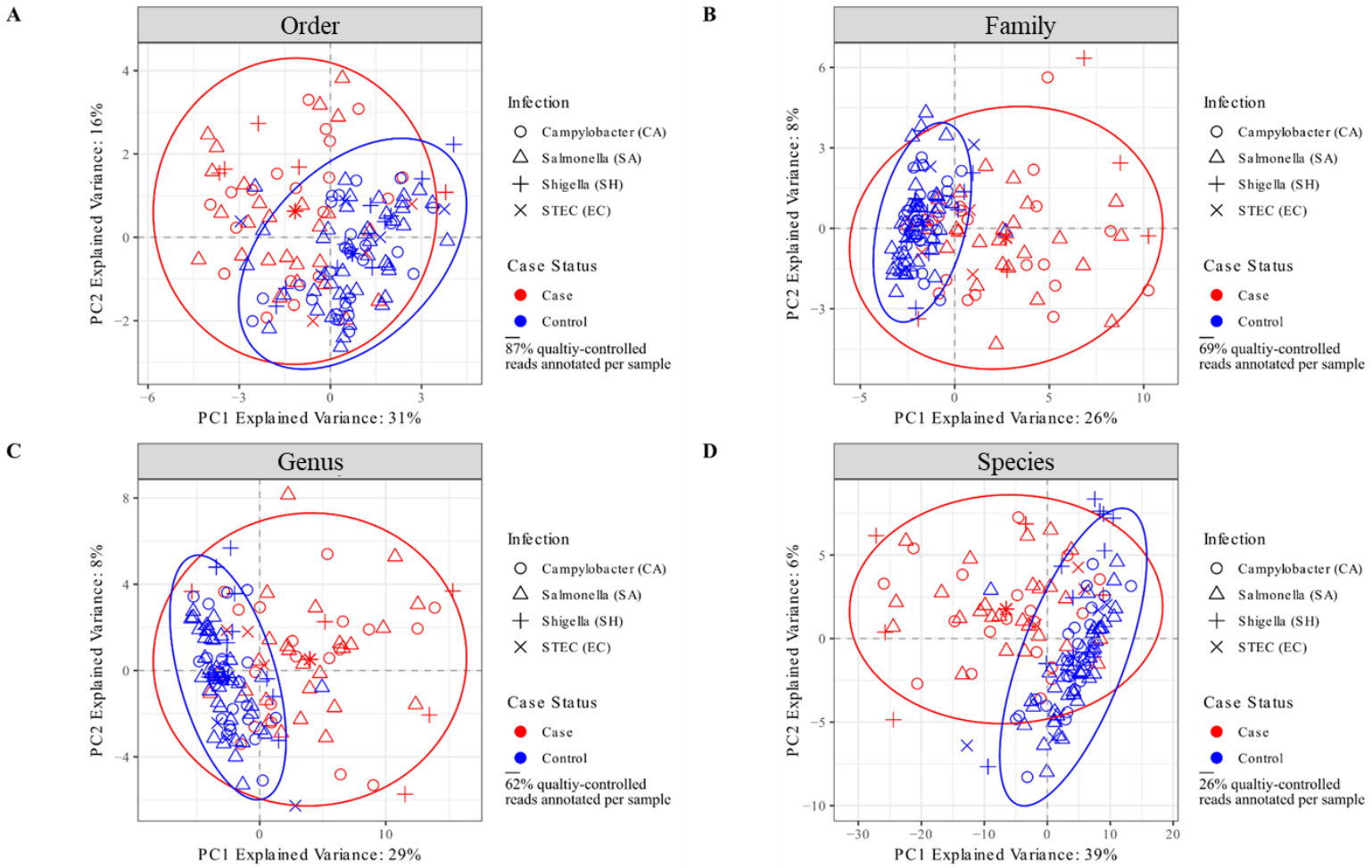


Figure 2.5. Rarefaction curves to evaluate the quality of sequencing. A) Random sampling was used to assess cumulative sequencing across all samples by study group, or cases (red line) versus controls (blue line); and B) Rarefaction to assess genera richness based on total reads sequenced across case (red line) and control (blue line) samples. The 95% confidence intervals are indicated for each curve.

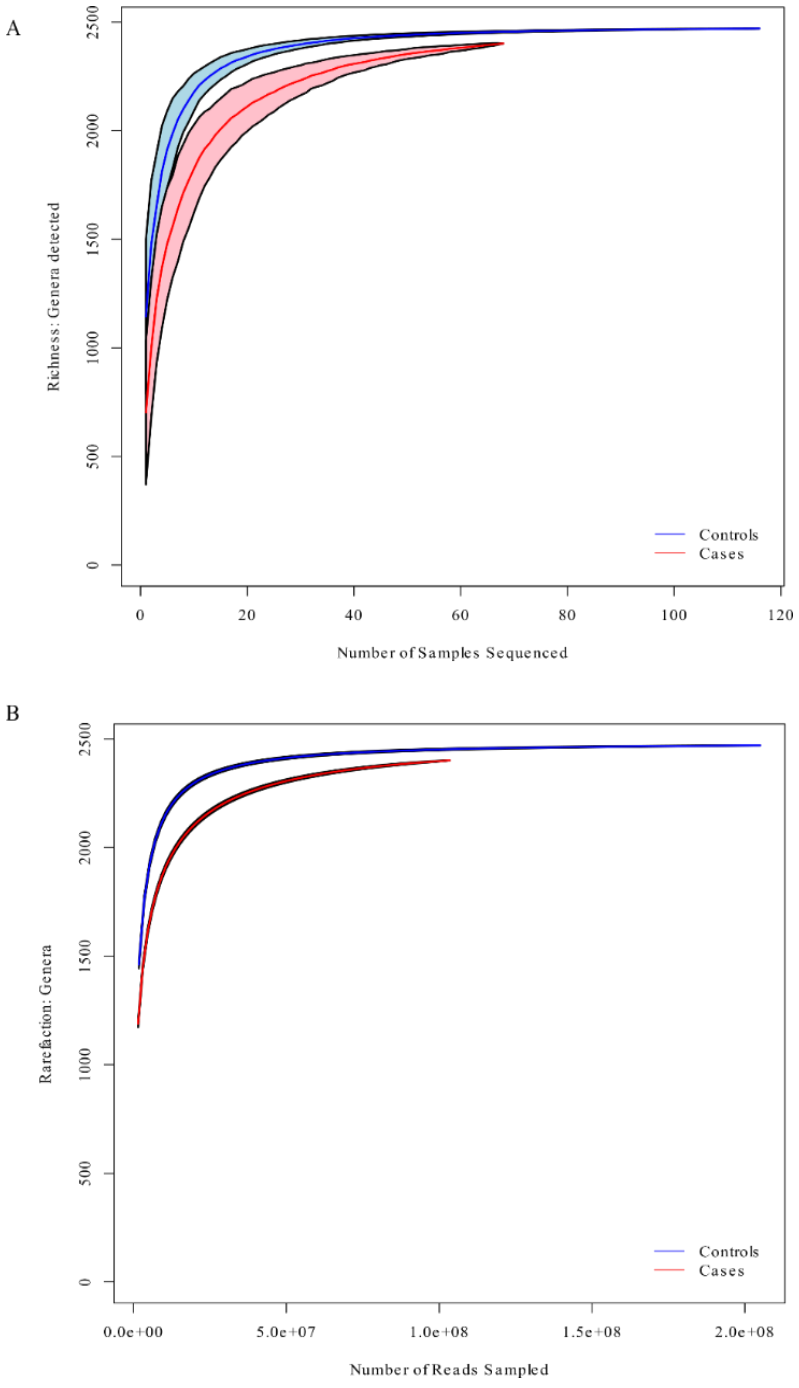


Figure 2.6. Metrics for case vs control to assess diversity A) Shannon Index determined by using the *diversity* function in R, B) Richness, the total number of species, as determined using the *specnumber* function in R C) Evenness, the distribution of species across each sample. Boxplots were generated for case and control groups: The line in the box represents the median. The interquartile range (25%-75%) is the box surrounding the median. The whiskers extend 1.5 times the interquartile range. Outliers are circles. Calculations performed using the total microbiome (bacteria and virus) at the genera level. * statistical significance ($p < 0.05$).

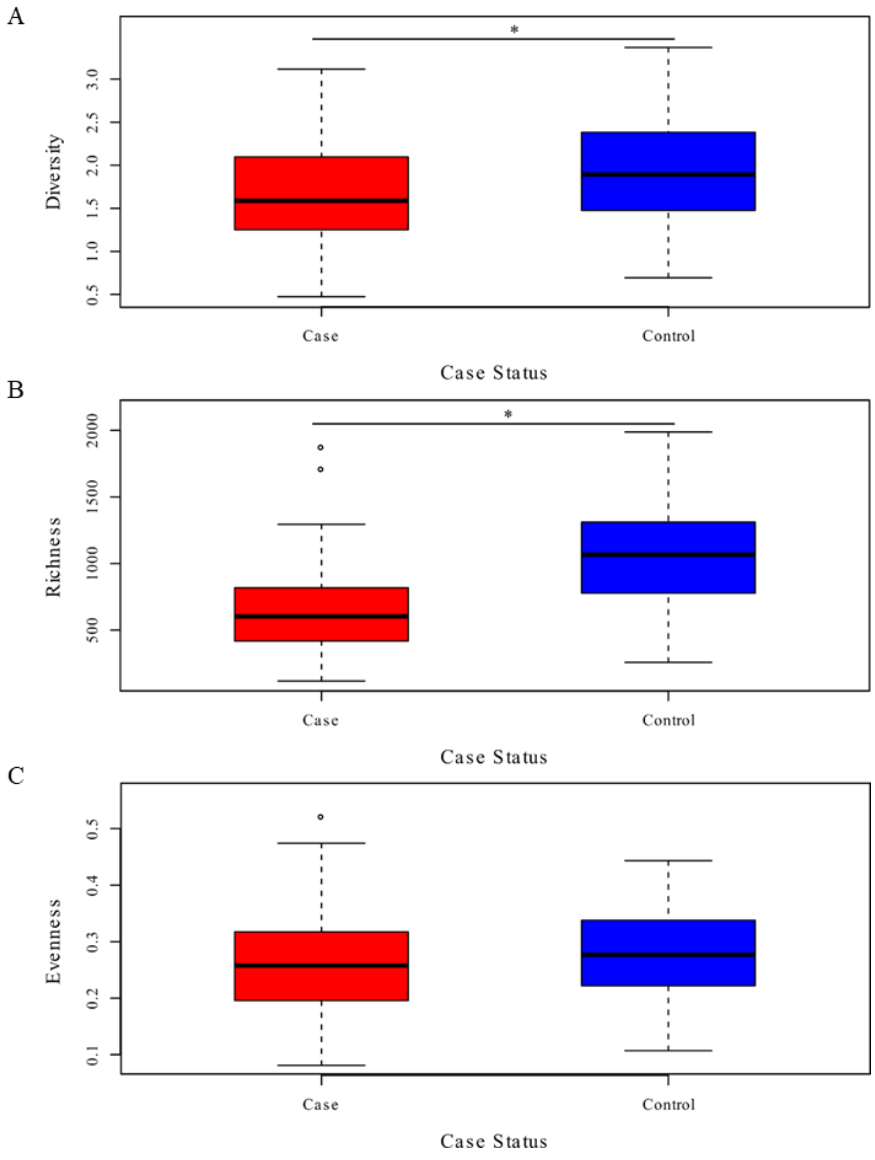


Figure 2.7. Microbiome profiles of Case and Control samples A) The top 5 highest abundant viruses across the study. B) The top 10 highest abundant viruses across the study. Both Viruses and Bacteria are presented at the Family taxonomical rank. The line in the box represents the median, the. The interquartile range (25%-75%) is the box surrounding the median. The whiskers extend 1.5 times the interquartile range. Outliers are circles.

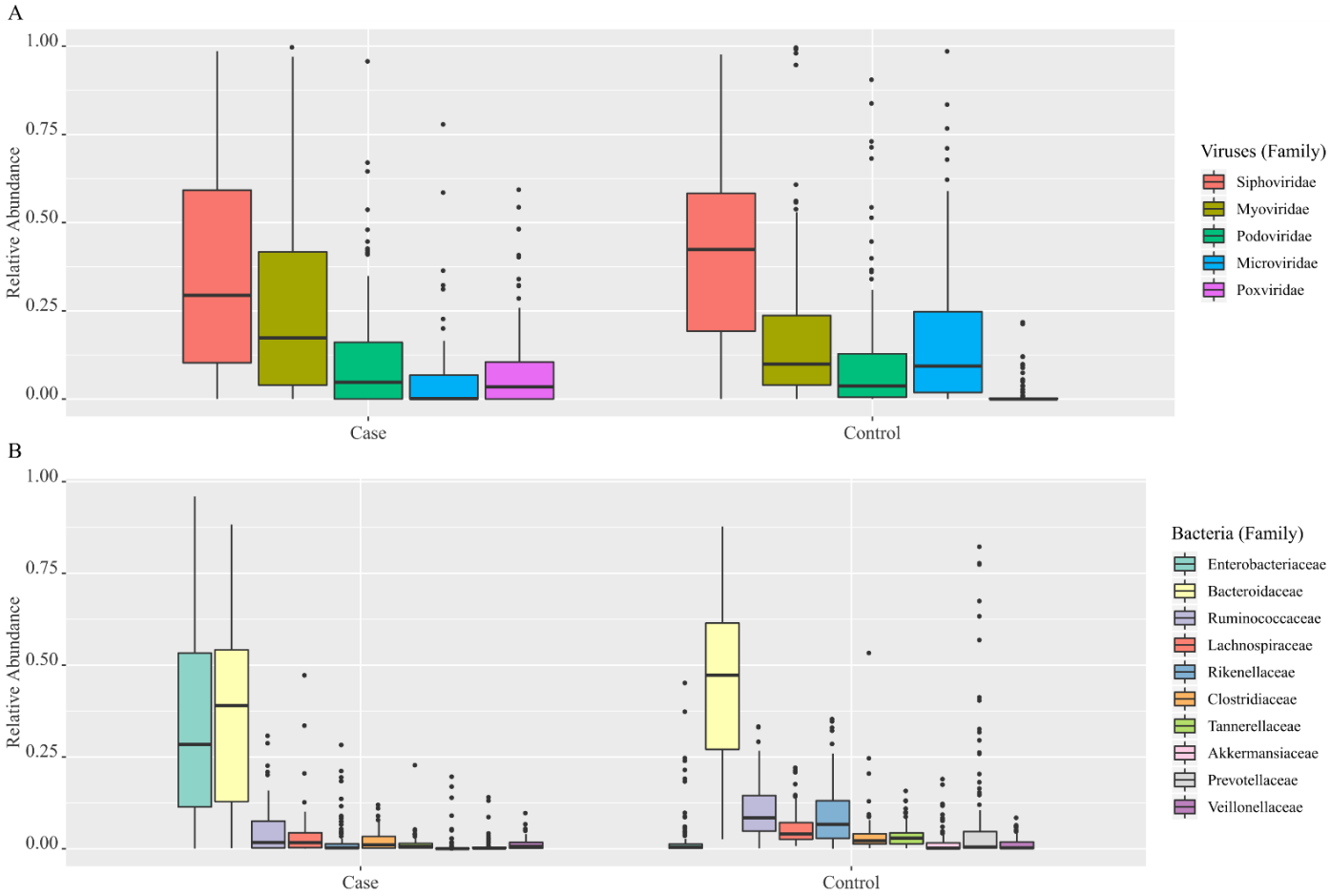


Figure 2.8. Microbiome profiles of Case by Infection type A) The top 10 most important virus by infection type. B) The top 10 most important bacteria by infection type across the study. Both Viruses and Bacteria are presented at the Genus taxonomical rank. The line in the box represents the median, the. The interquartile range (25%-75%) is the box surrounding the median. The whiskers extend from 1.5 times the interquartile range of samples. Outliers are circles.

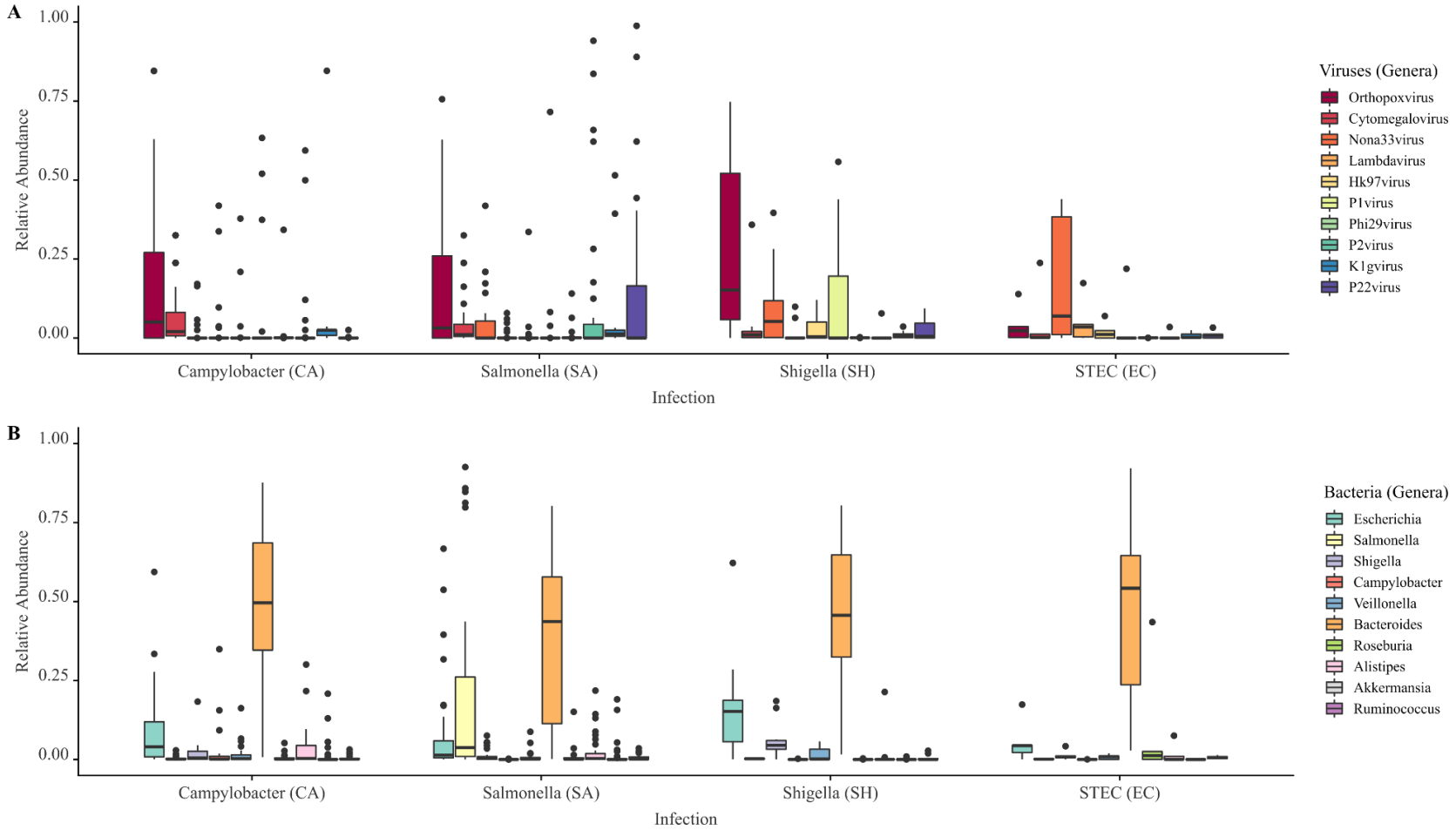


Figure 2.9. Distinct microbiome profiles identified by hierarchical clustering A) Four distinct clusters were identified using principal components analysis (PCA), and B) the beta dispersion of each cluster shows the spatial relationship of each sample within the cluster. For both panels, the legend, axes, and colors for each cluster are the same, while ellipses indicate the confidence interval (95%).

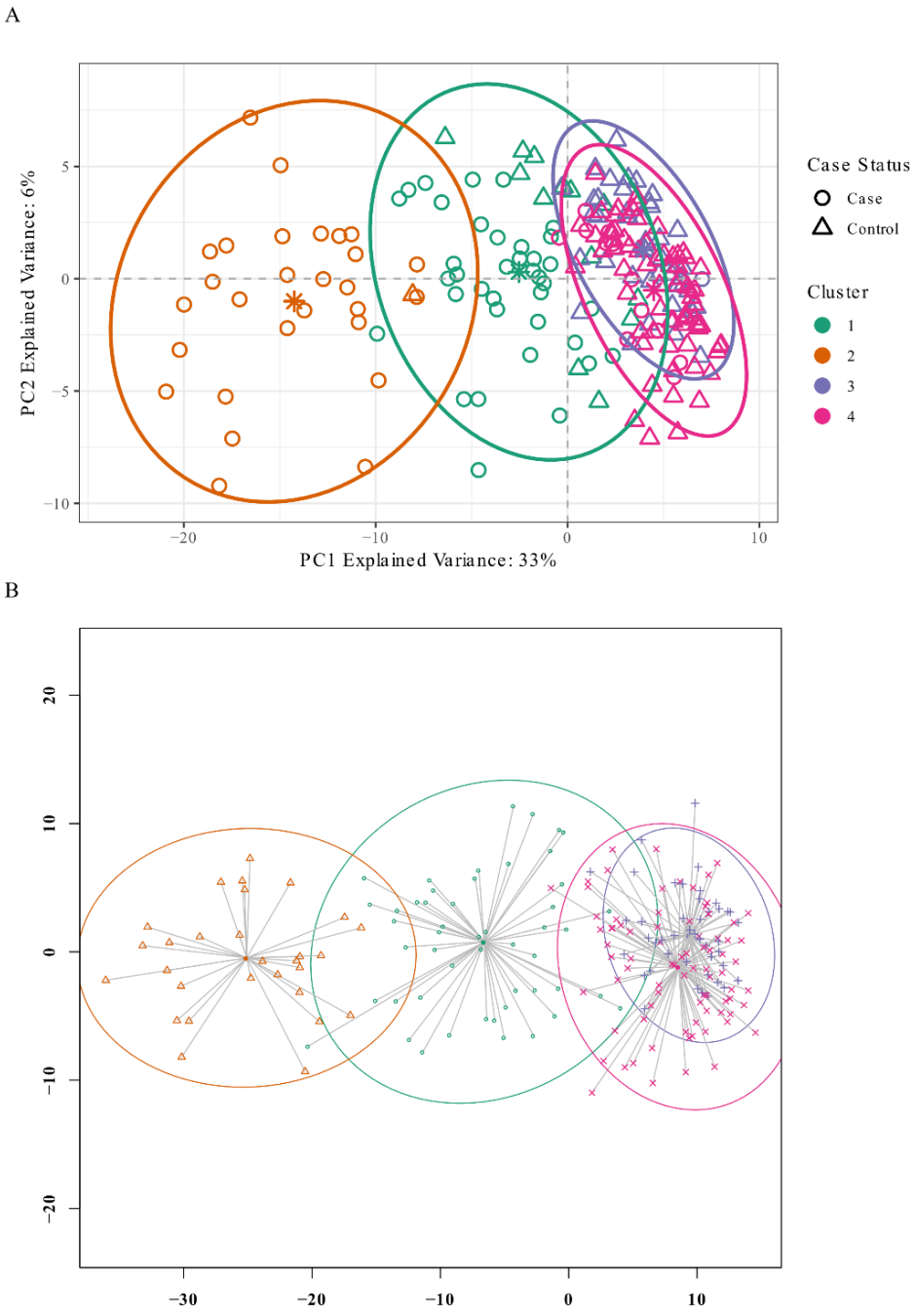


Figure 2.10. Metrics for clusters to assess diversity A) Shannon Diversity determined by using the *diversity* function in R, B) Richness, the total number of species was determined using the *specnumber* function in R C) Evenness which is the distribution of species across each sample. Boxplots were generated for each cluster: The line in the box represents the median. The interquartile range (25%-75%) is the box surrounding the median. The whiskers extend 1.5 times the interquartile range. Outliers are circles. Calculations performed using the total microbiome (bacteria and virus) at the genera level. * denotes statistical significance ($p < 0.05$).

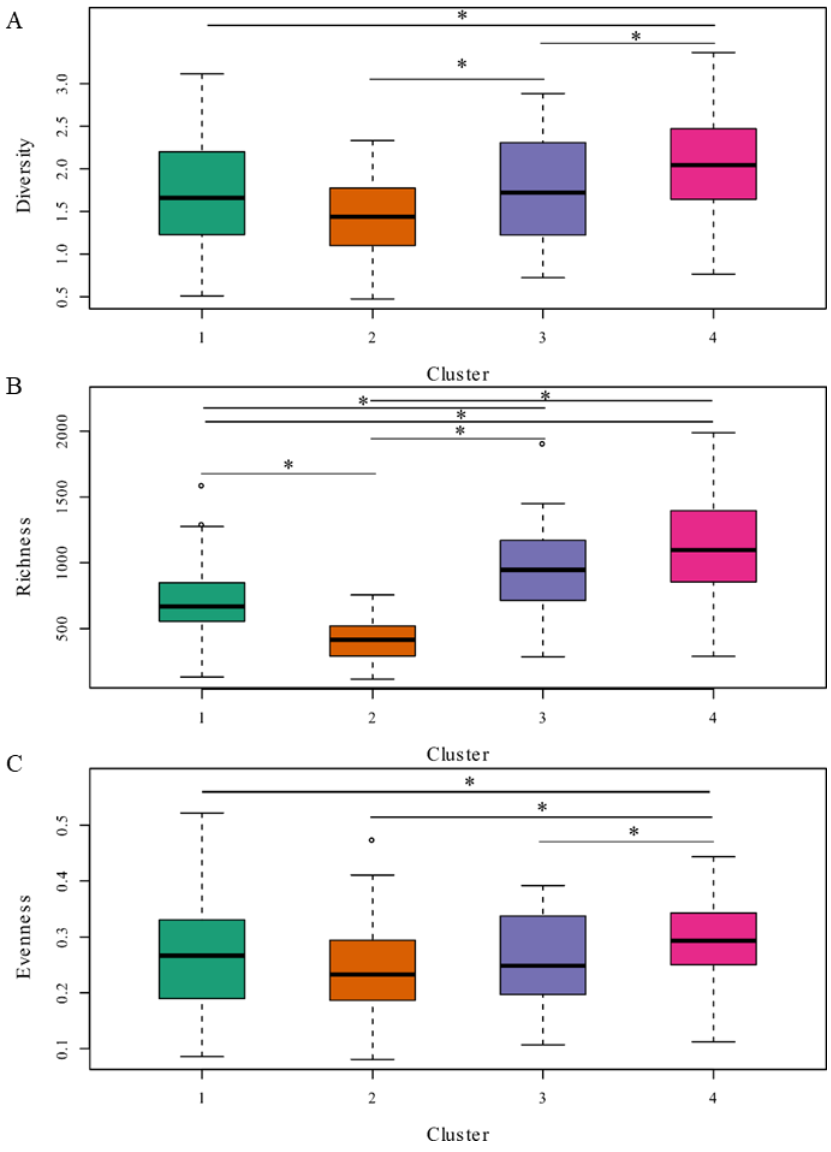


Figure 2.11. The four clusters have distinct microbiomes Clusters are colored as follows: Cluster 1 = green, Cluster 2 = orange, Cluster 3 = purple, Cluster 4 = pink. The coloring of the heatmap represents the Z-score or standard deviations from the mean within a column. Columns represent individual samples, and rows are taxa. Purple coloring represents more abundant taxa within a sample, whereas orange coloring represents lower abundant taxa. The dendrogram on the left represents the genera. Viruses are clustered at the bottom of the tree in V.

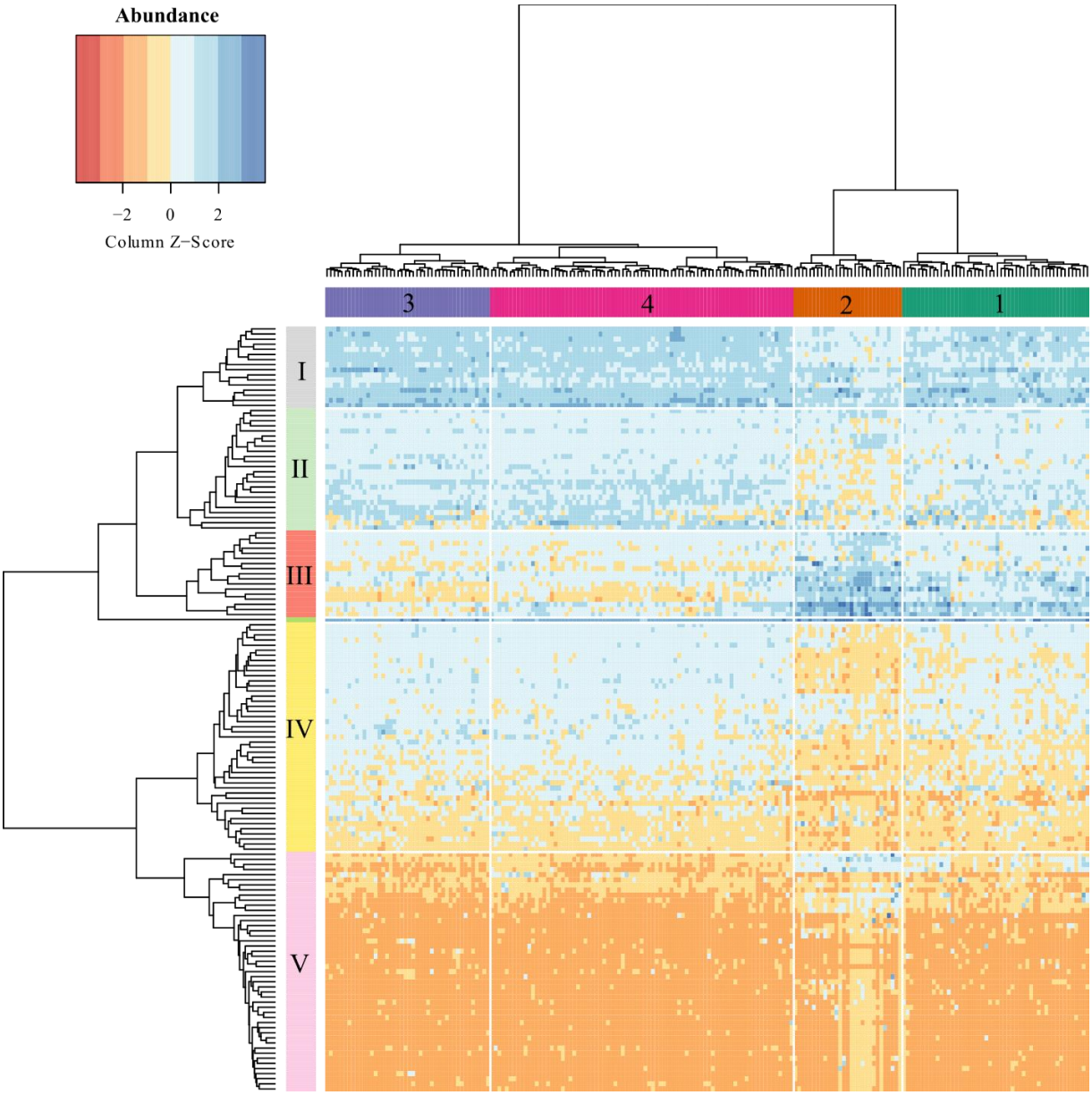


Figure 2.12. Case clusters have a common microbiome based on an analysis of 79 patients with enteric infections and 125 non-infected family members (controls) included in the study

A) Venn diagram showing the number of differentially abundant and shared taxa across all four clusters. B) Case clusters (Cluster 1 and Cluster 2) showing the number of differentially abundant and shared taxa. C) The most differentially abundant viruses across clusters D) The most differentially abundant bacteria across clusters.

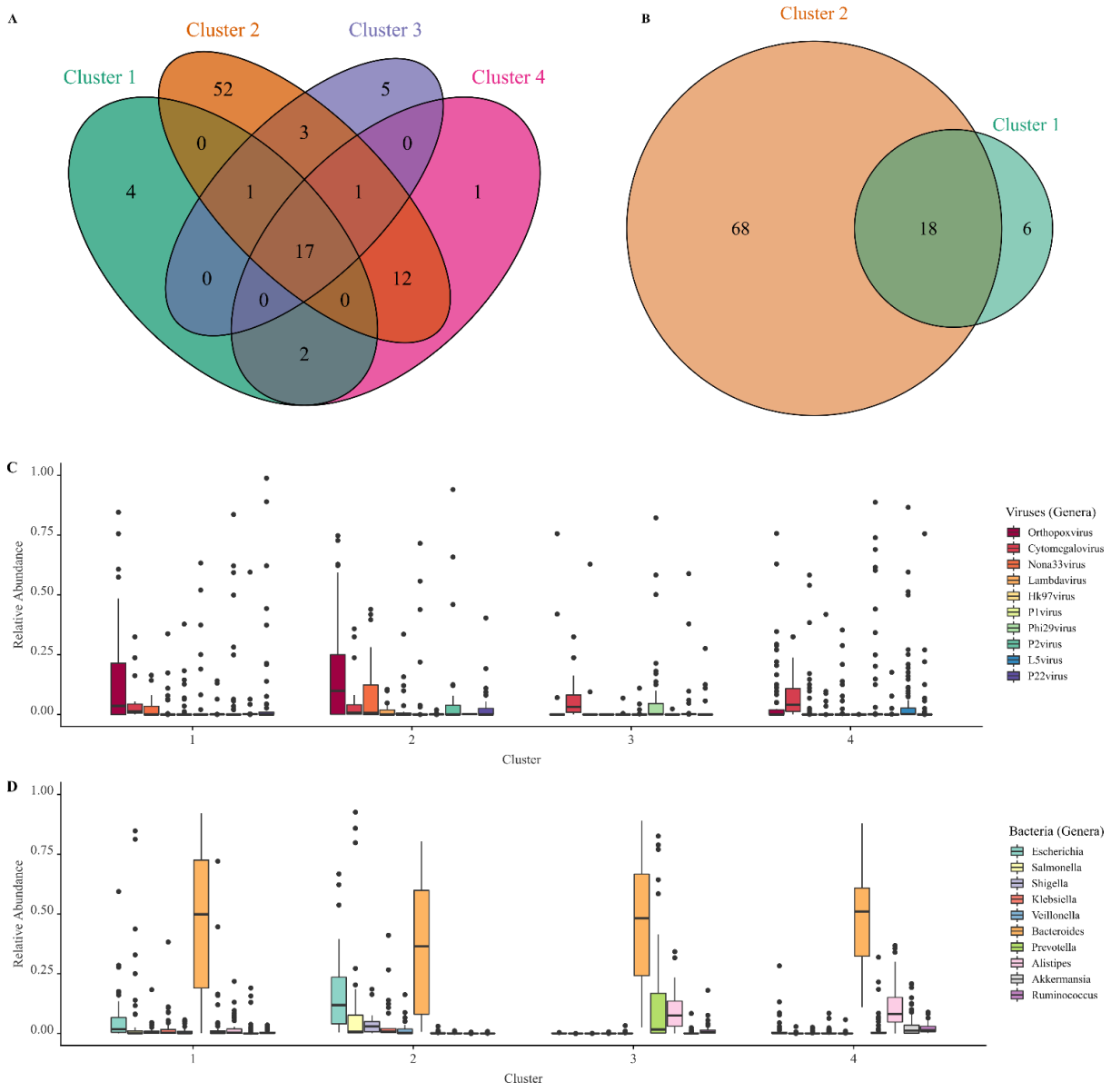
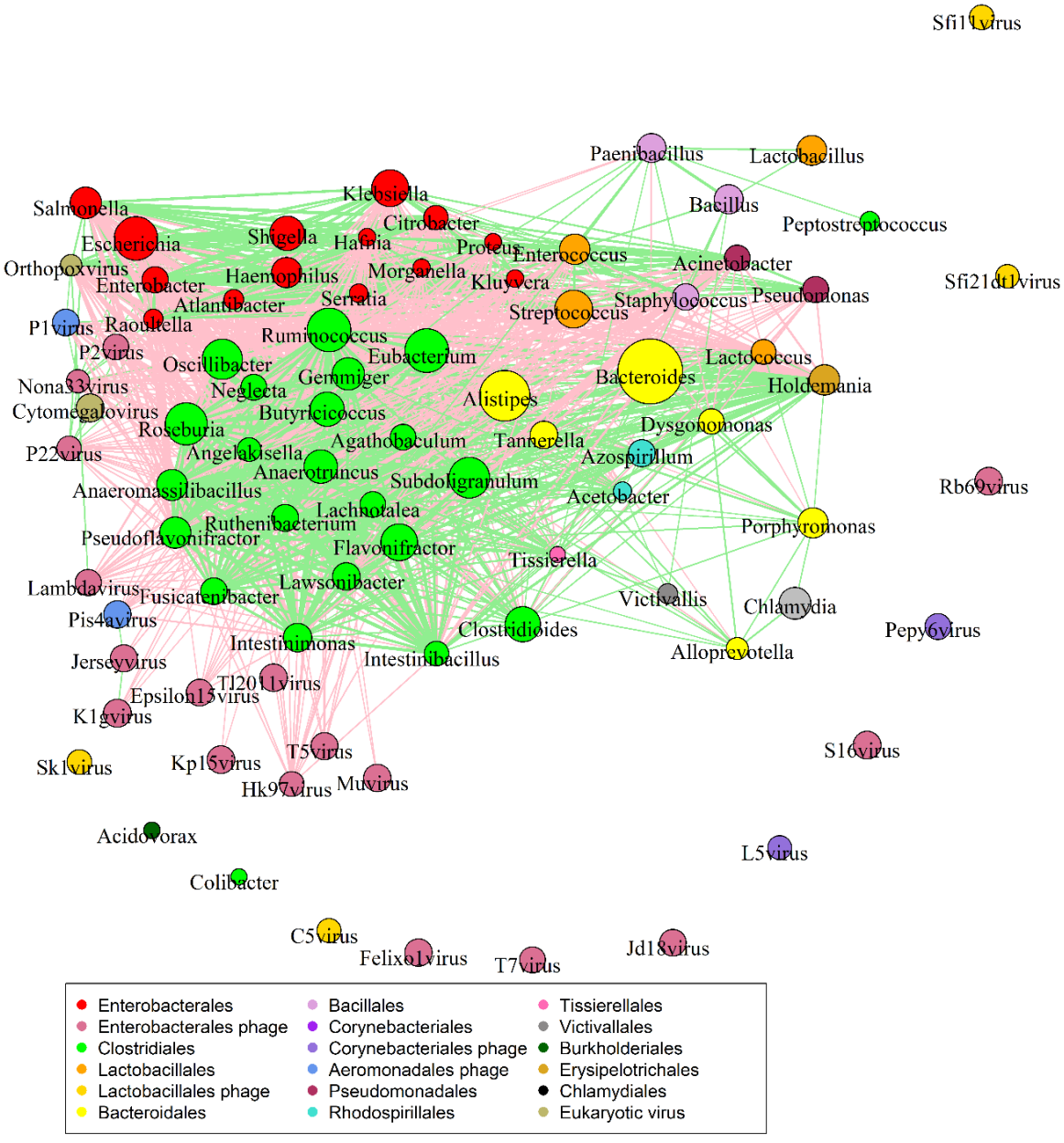


Figure 2.13. Network analysis of the microbes differentially abundant for Cluster 2 Sparcc with the Spieceasi pipeline was utilized to calculate correlations between taxa across samples. Edges represent correlations between taxa; positive correlations are in green, and negative correlations are red. The size of vertices represents the abundance found across samples and are colored based on higher taxonomical classification. Only significant correlations are represented (absolute value $\geq$ 0.3).



REFERENCES

REFERENCES

1. **Hall AJ**, **Rosenthal M**, **Gregoricus N**, **Greene SA**, **Ferguson J**, **Henao OL**, **Vinjé J**, **Lopman BA**, **Parashar UD**, **Widdowson MA**. 2011. Incidence of acute gastroenteritis and role of norovirus, Georgia, USA, 2004-2005. Emerg Infect Dis **17**:1381–1388.

2. **Herikstad H**, **Yang S**, **Van Gilder TJ**, **Vugia D**, **Hadler J**, **Blake P**, **Deneen V**, **Shiferaw B**, **Angulo FJ**, **The FOODNET Working Group**. 2002. A population-based estimate of the burden of diarrhoeal illness in the United States: FoodNet, 1996–7. Epidemiol Infect **129**:9–17.

3. **American Academy of Family Physicians., Hartman S**, **Brown E**, **Loomis E**, **Russell HA**. 2019. Gastroenteritis in Children. Am Fam Physician **99**:159–165.

4. **Vos T**, **Allen C**, **Arora M**, **Barber RM**, **Bhutta ZA**, **Brown A**, **Carter A**, **Casey DC**, **Charlson FJ**, **Chen AZ**, **Coggeshall M**, **Cornaby L**, **Murray CJL**. 2016. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet **388**:1545–1602.

5. **Kotloff KL**. 2017. The Burden and Etiology of Diarrheal Illness in Developing Countries. Pediatr Clin North Am **64**:799–814.

6. **Kotloff KL**, **Nataro JP**, **Blackwelder WC**, **Nasrin D**, **Farag TH**, **Panchalingam S**, **Wu Y**, **Sow SO**, **Sur D**, **Breiman RF**, **Faruque ASG**, **Zaidi AKM**, **Saha D**, **Alonso PL**, **Tamboura B**, **Sanogo D**, **Onwuchekwa U**, **Manna B**, **Ramamurthy T**, **Kanungo S**, **Ochieng JB**, **Omore R**, **Oundo JO**, **Hossain A**, **Das SK**, **Ahmed S**, **Qureshi S**, **Quadri F**, **Adegbola RA**, **Antonio M**, **Hossain MJ**, **Akinsola A**, **Mandomando I**, **Nhampossa T**, **Acácio S**, **Biswas K**, **O'Reilly CE**, **Mintz ED**, **Berkeley LY**, **Muhsen K**, **Sommerfelt H**, **Robins-Browne RM**, **Levine MM**. 2013. 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. Lancet **382**:209–222.

7. **Fletcher SM**, **Stark D**, **Ellis J**. 2011. Prevalence of gastrointestinal pathogens in Sub-Saharan Africa: systematic review and meta-analysis. J Public health Res **2**:e30.

8. **Braun T**, **Di Segni A**, **BenShoshan M**, **Asaf R**, **Squires JE**, **Farage Barhom S**, **Glick Saar E**, **Cesarkas K**, **Smollan G**, **Weiss B**, **Amit S**, **Keller N**, **Haberman Y**. 2017. Fecal microbial characterization of hospitalized patients with suspected infectious diarrhea shows significant dysbiosis. Sci Rep **7**:1088.

9. **American Academy of Family Physicians., Hartman S**, **Brown E**, **Loomis E**, **Russell HA**. 2019. Gastroenteritis in Children. Am Fam Physician **99**:159–165.

10. **Castaño-Rodríguez N**, **Underwood AP**, **Merif J**, **Riordan SM**, **Rawlinson WD**, **Mitchell HM**, **Kaakoush NO**. 2018. Gut microbiome analysis identifies potential etiological factors in acute gastroenteritis. Infect Immun **86**:1–13.

11. **Singh P**, **Teal TK**, **Marsh TL**, **Tiedje JM**, **Mosci R**, **Jernigan K**, **Zell A**, **Newton DW**,

Salimnia H, **Lephart P**, **Sundin D**, **Khalife W**, **Britton RA**, **Rudrik JT**, **Manning SD**. 2015. Intestinal microbial communities associated with acute enteric infections and disease recovery. Microbiome **3**:45.

12. **Kimmitt PT**, **Harwood CR**, **Barer MR**. 2000. Toxin gene expression by shiga toxin-producing Escherichia coli: the role of antibiotics and the bacterial SOS response. Emerg Infect Dis **6**:458–65.

13. **Muniesa M**, **Recktenwald J**, **Bielaszewska M**, **Karch H**, **Schmidt H**. 2000. Characterization of a Shiga toxin 2e-converting bacteriophage from an Escherichia coli strain of human origin. Infect Immun **68**:4850–4855.

14. **Hendaus MA**, **Jomha FA**, **Alhammadi AH**. 2015. Virus-induced secondary bacterial infection: A concise review. Ther Clin Risk Manag **11**:1265–1271.

15. **Package T**, **Champely AS**, **Champely S**. 2009. Package ‘ pwr .’ October 1–21.

16. **Ihaka R**, **Gentleman R**. 1996. R: A Language for Data Analysis and Graphics. J Comput Graph Stat.

17. **Cohen J**. 1988. Statistical power analysis for the behavioral sciences. Stat Power Anal Behav Sci.

18. **Bolger a. M**, **Lohse M**, **Usadel B**. 2014. Trimmomatic: A flexible read trimming tool for Illumina NGS data. Bioinformatics **30**:2114–2120.

19. **Andrews S**. 2010. FastQC: A quality control tool for high throughput sequence data. babraham Bioinforma http://www.bioinformatics.babraham.ac.uk/projects/.

20. **Norman JM**, **Handley SA**, **Baldridge MT**, **Droit L**, **Liu CY**, **Keller BC**, **Kambal A**, **Monaco CL**, **Zhao G**, **Fleshner P**, **Stappenbeck TS**, **McGovern DPB**, **Keshavarzian A**, **Mutlu EA**, **Sauk J**, **Gevers D**, **Xavier RJ**, **Wang D**, **Parkes M**, **Virgin HW**, **Norman JM**, **Handley SA**, **Baldridge MT**, **Droit L**, **Liu CY**, **Keller BC**. 2015. Disease-Specific Alterations in the Enteric Virome in Inflammatory Bowel Disease. Cell **160**:447–460.

21. **Cottingham RW**. 2015. The DOE systems biology knowledgebase (KBase).

22. **Langmead B**, **Salzberg SL**. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods **9**:357–9.

23. **Li H**, **Handsaker B**, **Wysoker A**, **Fennell T**, **Ruan J**, **Homer N**, **Marth G**, **Abecasis G**, **Durbin R**. 2009. The Sequence Alignment / Map (SAM) Format and SAMtools 1000 Genome Project Data Processing Subgroup. Bioinformatics **25**:1–2.

24. **Menzel P**, **Ng KL**, **Krogh A**. 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun **7**:11257.

25. **Agarwala R**, **Barrett T**, **Beck J**, **Benson DA**, **Bollin C**, **Bolton E**, **Bourexis D**, **Brister JR**, **Bryant SH**, **Canese K**, **Cavanaugh M**, **Charowhas C**, **Clark K**, **Dondoshansky I**, **Feolo M**, **Fitzpatrick L**, **Funk K**, **Geer LY**, **Gorelenkov V**, **Graeff A**, **Hlavina W**, **Holmes B**, **Johnson M**, **Kattman B**, **Khotomlianski V**, **Kimchi A**, **Kimelman M**, **Kimura M**, **Kitts P**, **Klimke W**, **Kotliarov A**, **Krasnov S**, **Kuznetsov A**, **Landrum MJ**,

Landsman D, **Lathrop S**, **Lee JM**, **Leubsdorf C**, **Lu Z**, **Madden TL**, **Marchler-Bauer A**, **Malheiro A**, **Meric P**, **Karsch-Mizrachi I**, **Mnev A**, **Murphy T**, **Orris R**, **Ostell J**, **O'Sullivan C**, **Palanigobu V**, **Panchenko AR**, **Phan L**, **Pierov B**, **Pruitt KD**, **Rodarmer K**, **Sayers EW**, **Schneider V**, **Schoch CL**, **Schuler GD**, **Sherry ST**, **Siyan K**, **Soboleva A**, **Soussov V**, **Starchenko G**, **Tatusova TA**, **Thibaud-Nissen F**, **Todorov K**, **Trawick BW**, **Vakatov D**, **Ward M**, **Yaschenko E**, **Zasypkin A**, **Zbicz K**. 2018. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res.

26. **Rossum G Van**. 2018. The Python / C API. Python Software,

27. **Arumugam M**, **Raes J**, **Pelletier E**, **Le Paslier D**, **Yamada T**, **Mende DR**, **Fernandes GR**, **Tap J**, **Bruls T**, **Batto JM**, **Bertalan M**, **Borruel N**, **Casellas F**, **Fernandez L**, **Gautier L**, **Hansen T**, **Hattori M**, **Hayashi T**, **Kleerebezem M**, **Kurokawa K**, **Leclerc M**, **Levenez F**, **Manichanh C**, **Nielsen HB**, **Nielsen T**, **Pons N**, **Poulain J**, **Qin J**, **Sicheritz-Ponten T**, **Tims S**, **Torrents D**, **Ugarte E**, **Zoetendal EG**, **Wang J**, **Guarner F**, **Pedersen O**, **De Vos WM**, **Brunak S**, **Doré J**, **Weissenbach J**, **Ehrlich SD**, **Bork P**. 2011. Enterotypes of the human gut microbiome. Nature.

28. **Nurk S**, **Meleshko D**, **Korobeynikov A**, **Pevzner PA**. 2017. metaSPAdes: a new versatile metagenomic assembler (Supplementary Material). Genome Res **27**:824–834.

29. **Navas-Molina JA**, **Peralta-Sánchez JM**, **González A**, **McMurdie PJ**, **Vázquez-Baeza Y**, **Xu Z**, **Ursell LK**, **Lauber C**, **Zhou H**, **Song SJ**, **Huntley J**, **Ackermann GL**, **Berg-Lyons D**, **Holmes S**, **Caporaso JG**, **Knight R**. 2013. Advancing our understanding of the human microbiome using QIIMEMethods in Enzymology.

30. **Hughes JB**, **Hellmann JJ**. 2005. The application of rarefaction techniques to molecular inventories of microbial diversity. Methods Enzymol.

31. **McMurdie PJ**, **Holmes S**. 2014. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. PLoS Comput Biol.

32. **Sanders HL**. 1968. Marine Benthic Diversity: A Comparative Study. Am Nat.

33. **Hurlbert SH**. 1971. The Nonconcept of Species Diversity: A Critique and Alternative Parameters. Ecology.

34. **Heck KL**, **van Belle G**, **Simberloff D**. 1975. Explicit Calculation of the Rarefaction Diversity Measurement and the Determination of Sufficient Sample Size. Ecology.

35. **Codoñer FM**, **Ramírez-Bosca A**, **Climent E**, **Carrión-Gutierrez M**, **Guerrero M**, **Pérez-Orquín JM**, **Horga De La Parte J**, **Genovés S**, **Ramón D**, **Navarro-López V**, **Chenoll E**. 2018. Gut microbial composition in patients with psoriasis. Sci Rep.

36. **Rodriguez-R LM**, **Gunturu S**, **Tiedje JM**, **Cole JR**, **Konstantinidis KT**. 2018. Nonpareil 3: Fast Estimation of Metagenomic Coverage and Sequence Diversity. mSystems.

37. **Rohart F**, **Gautier B**, **Singh A**, **Lê Cao KA**. 2017. mixOmics: An R package for 'omics feature selection and multiple data integration. PLoS Comput Biol.

38. **Palarea-Albaladejo J**, **Martín-Fernández JA**. 2015. ZCompositions - R package for multivariate imputation of left-censored data under a compositional approach. Chemom Intell Lab Syst.

39. **R Core Team**. 2017. R: A language and environment for statistical computing. http://www.R-project.org/. R Found Stat Comput Vienna, Austria.

40. **Pärnänen K**, **Karkman A**, **Hultman J**, **Lyra C**, **Bengtsson-Palme J**, **Larsson DGJ**, **Rautava S**, **Isolauri E**, **Salminen S**, **Kumar H**, **Satokari R**, **Virta M**. 2018. Maternal gut and breast milk microbiota affect infant gut antibiotic resistome and mobile genetic elements. Nat Commun.

41. **Gloor GB**, **Macklaim JM**, **Pawlowsky-Glahn V**, **Egozcue JJ**. 2017. Microbiome datasets are compositional: And this is not optional. Front Microbiol.

42. **Aitchison J**. 1982. The Statistical Analysis of Compositional Data. J R Stat Soc Ser B.

43. **Jari Oksanen, F. Guillaume Blanchet, Michael Friendly RK**, **Pierre Legendre, Dan McGlinn, Peter R. Minchin RBO**, **Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens ES**, **Helene Wagne**. 2017. Package 'vegan' | Community Ecology Package 1–292.

44. **Mandal S**, **Treuren W Van**, **White RA**, **Eggesbø M**, **Knight R**, **Peddada SD**. 2015. Analysis of composition of microbiomes: a novel method for studying microbial composition **1**:1–7.

45. **Kurtz ZD**, **Müller CL**, **Miraldi ER**, **Littman DR**, **Blaser MJ**, **Bonneau RA**. 2015. Sparse and Compositionally Robust Inference of Microbial Ecological Networks. PLoS Comput Biol.

46. **Centers for Disease Control and Prevention**. 2015. Introduction to Epi Info ™ 7 Using Epi Info ™. Epi Info.

47. **Hosmer DW**, **Lemeshow S**. 1980. Goodness of fit tests for the multiple logistic regression model. Commun Stat - Theory Methods.

48. **Tessler M**, **Neumann JS**, **Afshinnekoo E**, **Pineda M**, **Hersch R**, **Velho LFM**, **Segovia BT**, **Lansac-Toha FA**, **Lemke M**, **Desalle R**, **Mason CE**, **Brugler MR**. 2017. Large-scale differences in microbial biodiversity discovery between 16S amplicon and shotgun sequencing. Sci Rep.

49. **Nelson MC**, **Morrison HG**, **Benjamino J**, **Grim SL**, **Graf J**. 2014. Analysis, optimization and verification of illumina-generated 16s rRNA gene amplicon surveys. PLoS One.

50. **Guo J**, **Cole JR**, **Zhang Q**, **Brown T**, **Tiedje JM**. 2016. Microbial community analysis with ribosomal gene fragments from shotgun metagenomes. Appl Environ Microbiol.

51. **Logares R**, **Sunagawa S**, **Salazar G**, **Cornejo-Castillo FM**, **Ferrera I**, **Sarmento H**, **Hingamp P**, **Ogata H**, **de Vargas C**, **Lima-Mendez G**, **Raes J**, **Poulain J**, **Jaillon O**, **Wincker P**, **Kandels-Lewis S**, **Karsenti E**, **Bork P**, **Acinas SG**. 2014. Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity

and structure of microbial communities. Environ Microbiol.

52. **Shakya M**, **Quince C**, **Campbell JH**, **Yang ZK**, **Schadt CW**, **Podar M**. 2013. Comparative metagenomic and rRNA microbial diversity characterization using archaeal and bacterial synthetic communities. Environ Microbiol.

53. **Segata N**, **Waldron L**, **Ballarini A**, **Narasimhan V**, **Jousson O**, **Huttenhower C**. 2012. Metagenomic microbial community profiling using unique clade-specific marker genes. Nat Methods.

54. **Sender R**, **Fuchs S**, **Milo R**. 2016. Revised Estimates for the Number of Human and Bacteria Cells in the Body. PLoS Biol.

55. **Dutilh BE**, **Cassman N**, **McNair K**, **Sanchez SE**, **Silva GGZ**, **Boling L**, **Barr JJ**, **Speth DR**, **Seguritan V**, **Aziz RK**, **Felts B**, **Dinsdale EA**, **Mokili JL**, **Edwards RA**. 2014. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. Nat Commun.

56. **Zaheer R**, **Noyes N**, **Ortega Polo R**, **Cook SR**, **Marinier E**, **Van Domselaar G**, **Belk KE**, **Morley PS**, **McAllister TA**. 2018. Impact of sequencing depth on the characterization of the microbiome and resistome. Sci Rep.

57. **Xia Y**, **Sun J**, **Chen D-G**. 2018. Multivariate Community Analysis.

58. **Vincent C**, **Mehrotra S**, **Loo VG**, **Dewar K**, **Manges AR**. 2015. Excretion of Host DNA in Feces Is Associated with Risk of Clostridium difficile Infection. J Immunol Res.

59. **Lewis JD**, **Chen EZ**, **Baldassano RN**, **Otley AR**, **Griffiths AM**, **Lee D**, **Bittinger K**, **Bailey A**, **Friedman ES**, **Hoffmann C**, **Albenberg L**, **Sinha R**, **Compher C**, **Gilroy E**, **Nessel L**, **Grant A**, **Chehoud C**, **Li H**, **Wu GD**, **Bushman FD**. 2015. Inflammation, Antibiotics, and Diet as Environmental Stressors of the Gut Microbiome in Pediatric Crohn's Disease. Cell Host Microbe.

60. **Klaassen CHW**, **Jeunink MAF**, **Prinsen CFM**, **Ruers TJM**, **Tan ACITL**, **Strobbe LJA**, **Thunnissen FBJM**. 2003. Quantification of human DNA in feces as a diagnostic test for the presence of colorectal cancer. Clin Chem.

61. **Nairz M**, **Schroll A**, **Sonnweber T**, **Weiss G**. 2010. The struggle for iron - a metal at the host-pathogen interface. Cell Microbiol.

62. **Shin NR**, **Whon TW**, **Bae JW**. 2015. Proteobacteria: Microbial signature of dysbiosis in gut microbiota. Trends Biotechnol.

63. **Monaco CL**, **Gootenberg DB**, **Zhao G**, **Handley SA**, **Ghebremichael MS**, **Lim ES**, **Lankowski A**, **Baldridge MT**, **Wilen CB**, **Flagg M**, **Norman JM**, **Keller BC**, **Lu??vano JM**, **Wang D**, **Boum Y**, **Martin JN**, **Hunt PW**, **Bangsberg DR**, **Siedner MJ**, **Kwon DS**, **Virgin HW**. 2016. Altered Virome and Bacterial Microbiome in Human Immunodeficiency Virus-Associated Acquired Immunodeficiency Syndrome. Cell Host Microbe **19**:311–322.

64. **Nelson AM**, **Walk ST**, **Taube S**, **Taniuchi M**, **Houpt ER**, **Wobus CE**, **Young VB**. 2012. Disruption of the Human Gut Microbiota following Norovirus Infection. PLoS One

7.

65. **Saulnier DM**, **Riehle K**, **Mistretta TA**, **Diaz MA**, **Mandal D**, **Raza S**, **Weidler EM**, **Qin X**, **Coarfa C**, **Milosavljevic A**, **Petrosino JF**, **Highlander S**, **Gibbs R**, **Lynch S V.**, **Shulman RJ**, **Versalovic J**. 2011. Gastrointestinal microbiome signatures of pediatric patients with irritable bowel syndrome. Gastroenterology.

66. **Youmans BP**, **Ajami NJ**, **Jiang Z-D**, **Campbell F**, **Wadsworth WD**, **Petrosino JF**, **DuPont HL**, **Highlander SK**. 2015. Characterization of the human gut microbiome during travelers' diarrhea. Gut Microbes **6**:110–119.

67. **Breitbart M**, **Hewson I**, **Felts B**, **Mahaffy JM**, **Nulton J**, **Salamon P**, **Rohwer F**. 2003. Metagenomic analyses of an uncultured viral community from human feces. J Bacteriol **185**:6220–6223.

68. **Reyes A**, **Blanton L V.**, **Cao S**, **Zhao G**, **Manary M**, **Trehan I**, **Smith MI**, **Wang D**, **Virgin HW**, **Rohwer F**, **Gordon JI**. 2015. Gut DNA viromes of Malawian twins discordant for severe acute malnutrition. Proc Natl Acad Sci U S A **112**:11941–11946.

69. **Lim ES**, **Zhou Y**, **Zhao G**, **Bauer IK**, **Droit L**, **Ndao IM**, **Warner BB**, **Tarr PI**, **Wang D**, **Holtz LR**. 2015. Early life dynamics of the human gut virome and bacterial microbiome in infants. Nat Med **21**:1228–34.

70. **Owen S V.**, **Wenner N**, **Canals R**, **Makumi A**, **Hammarlöf DL**, **Gordon MA**, **Aertsen A**, **Feasey NA**, **Hinton JCD**. 2017. Characterization of the prophage repertoire of African Salmonella Typhimurium ST313 reveals high levels of spontaneous induction of novel phage BTP1. Front Microbiol.

71. **Adams MJ**, **Lefkowitz EJ**, **King AMQ**, **Harrach B**, **Harrison RL**, **Knowles NJ**, **Kropinski AM**, **Krupovic M**, **Kuhn JH**, **Mushegian AR**, **Nibert M**, **Sabanadzovic S**, **Sanfaçon H**, **Siddell SG**, **Simmonds P**, **Varsani A**, **Zerbini FM**, **Gorbalenya AE**, **Davison AJ**. 2016. Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses (2016). Arch Virol.

72. **Duerkop B a**, **Clements C V**, **Rollins D**, **Rodrigues JLM**, **Hooper L V**. 2012. A composite bacteriophage alters colonization by an intestinal commensal bacterium. Proc Natl Acad Sci U S A **109**:17621–6.

73. **Thingstad T**. 1998. A theorethical approach to structuring mechanisms in the pelagial food web. Hydrobiol **3631359**:59–72.

74. **Bohannan BJM**, **Lenski RE**. 1997. Effect of resource enrichment on a chemostat community of bacteria and bacteriophage. Ecology **78**:2303–2315.

75. **Rodriguez-Brito B**, **Li L**, **Wegley L**, **Furlan M**, **Angly F**, **Breitbart M**, **Buchanan J**, **Desnues C**, **Dinsdale E**, **Edwards R**, **Felts B**, **Haynes M**, **Liu H**, **Lipson D**, **Mahaffy J**, **Martin-Cuadrado AB**, **Mira A**, **Nulton J**, **Pašić L**, **Rayhawk S**, **Rodriguez-Mueller J**, **Rodriguez-Valera F**, **Salamon P**, **Srinagesh S**, **Thingstad TF**, **Tran T**, **Thurber RV**, **Willner D**, **Youle M**, **Rohwer F**. 2010. Viral and microbial community dynamics in four aquatic environments. ISME J **4**:739–751.

76. **Winter C**, **Bouvier T**, **Weinbauer MG**, **Thingstad TF**. 2010. Trade-Offs between

Competition and Defense Specialists among Unicellular Planktonic Organisms: the "Killing the Winner" Hypothesis Revisited. Microbiol Mol Biol Rev **74**:42–57.

77. **Pride DT**, **Salzman J**, **Haynes M**, **Rohwer F**, **Davis-Long C**, **White 3rd RA**, **Loomer P**, **Armitage GC**, **Relman DA**. 2012. Evidence of a robust resident bacteriophage population revealed through analysis of the human salivary virome. Isme J **6**:915–926.

78. **Barr JJ**, **Auro R**, **Furlan M**, **Whiteson KL**, **Erb ML**, **Pogliano J**, **Stotland A**, **Wolkowicz R**, **Cutting AS**, **Doran KS**, **Salamon P**, **Youle M**, **Rohwer F**. 2013. Bacteriophage adhering to mucus provide a non-host-derived immunity. Proc Natl Acad Sci U S A **110**:10771–6.

79. **Díaz-Muñoz SL**, **Koskella B**. 2014. Bacteria–Phage Interactions in Natural Environments, p. 135–183. *In* Advances in Applied Microbiology.

80. **Gorvitovskaia A**, **Holmes SP**, **Huse SM**. 2016. Interpreting prevotella and bacteroides as biomarkers of diet and lifestyle. Microbiome.

81. **Chen S-Y**, **Tsai C-N**, **Lee Y-S**, **Lin C-Y**, **Huang K-Y**, **Chao H-C**, **Lai M-W**, **Chiu C-H**. 2017. Intestinal microbiome in children with severe and complicated acute viral gastroenteritis. Sci Rep **7**:46130.

82. **Van Herreweghen F**, **De Paepe K**, **Roume H**, **Kerckhof FM**, **Van de Wiele T**. 2018. Mucin degradation niche as a driver of microbiome composition and Akkermansia muciniphila abundance in a dynamic gut model is donor independent. FEMS Microbiol Ecol.

83. **Desai MS**, **Seekatz AM**, **Koropatkin NM**, **Kamada N**, **Hickey CA**, **Wolter M**, **Pudlo NA**, **Kitamoto S**, **Terrapon N**, **Muller A**, **Young VB**, **Henrissat B**, **Wilmes P**, **Stappenbeck TS**, **N????ez G**, **Martens EC**. 2016. A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. Cell **167**:1339-1353.e21.

84. **Göker M**, **Gronow S**, **Zeytun A**, **Nolan M**, **Lucas S**, **Lapidus A**, **Hammon N**, **Deshpande S**, **Cheng JF**, **Pitluck S**, **Liolios K**, **Pagani I**, **Ivanova N**, **Mavromatis K**, **Ovchinikova G**, **Pati A**, **Tapia R**, **Han C**, **Goodwin L**, **Chen A**, **Palaniappan K**, **Land M**, **Hauser L**, **Jeffries CD**, **Brambilla EM**, **Rohde M**, **Detter JC**, **Woyke T**, **Bristow J**, **Markowitz V**, **Hugenholtz P**, **Eisen JA**, **Kyrpides NC**, **Klenk HP**. 2011. Complete genome sequence of odoribacter splanchnicus type strain (1651/6 T). Stand Genomic Sci **4**:200–209.

85. **Wang Z**, **Wang Q**, **Zhao J**, **Gong L**, **Zhang Y**, **Wang X**, **Yuan Z**. 2019. Altered diversity and composition of the gut microbiome in patients with cervical cancer. AMB Express **9**.

86. **Goodrich JK**, **Waters JL**, **Poole AC**, **Sutter JL**, **Koren O**, **Blekhman R**, **Beaumont M**, **Van Treuren W**, **Knight R**, **Bell JT**, **Spector TD**, **Clark AG**, **Ley RE**. 2014. Human genetics shape the gut microbiome. Cell **159**:789–799.

87. **Duncan SH**, **Hold GL**, **Barcenilla A**, **Stewart CS**, **Flint HJ**. 2002. Roseburia intestinalis sp. nov., a novel saccharolytic, butyrate-producing bacterium from human faeces. Int J

Syst Evol Microbiol.

88. **Segain JP**, **Galmiche JP**, **Raingeard De La Blétière D**, **Bourreille A**, **Leray V**, **Gervois N**, **Rosales C**, **Ferrier L**, **Bonnet C**, **Blottière HM**. 2000. Butyrate inhibits inflammatory responses through NFκB inhibition: Implications for Crohn's disease. Gut.

89. **Cekanaviciute E**, **Yoo BB**, **Runia TF**, **Debelius JW**, **Singh S**, **Nelson CA**, **Kanner R**, **Bencosme Y**, **Lee YK**, **Hauser SL**, **Crabtree-Hartman E**, **Sand IK**, **Gacias M**, **Zhu Y**, **Casaccia P**, **Cree BAC**, **Knight R**, **Mazmanian SK**, **Baranzini SE**. 2017. Gut bacteria from multiple sclerosis patients modulate human T cells and exacerbate symptoms in mouse models. Proc Natl Acad Sci U S A.

90. **Ness IF**, **Diep DB**, **Ike Y**. 2014. Enterococcal Bacteriocins and Antimicrobial Proteins that Contribute to Niche ControlEnterococci: From Commensals to Leading Causes of Drug Resistant Infection.

91. **Gogokhia L**, **Buhrke K**, **Bell R**, **Hoffman B**, **Brown DG**, **Hanke-Gogokhia C**, **Ajami NJ**, **Wong MC**, **Ghazaryan A**, **Valentine JF**, **Porter N**, **Martens E**, **O'Connell R**, **Jacob V**, **Scherl E**, **Crawford C**, **Stephens WZ**, **Casjens SR**, **Longman RS**, **Round JL**. 2019. Expansion of Bacteriophages Is Linked to Aggravated Intestinal Inflammation and Colitis. Cell Host Microbe **25**:285-299.e8.

92. **Reyes A**, **Wu M**, **McNulty NP**, **Rohwer FL**, **Gordon JI**. 2013. Gnotobiotic mouse model of phage-bacterial host dynamics in the human gut. Proc Natl Acad Sci U S A **110**:20236–41.

93. **Alcami A**. 2003. Viral mimicry of cytokines, chemokines and their receptors. Nat Rev Immunol.

94. **Finlay BB**, **McFadden G**. 2006. Anti-immunology: Evasion of the host immune system by bacterial and viral pathogens. Cell.

95. **De Cárcer DA**, **Hernáez B**, **Rastrojo A**, **Alcamí A**. 2017. Infection with diverse immune-modulating poxviruses elicits different compositional shifts in the mouse gut microbiome. PLoS One **12**:1–9.

96. **Kernbauer E**, **Ding Y**, **Cadwell K**. 2014. An enteric virus can replace the beneficial function of commensal bacteria. Nature **516**:94–98.

97. **Yang JY**, **Kim MS**, **Kim E**, **Cheon JH**, **Lee YS**, **Kim Y**, **Lee SH**, **Seo SU**, **Shin SH**, **Choi SS**, **Kim B**, **Chang SY**, **Ko HJ**, **Bae JW**, **Kweon MN**. 2016. Enteric Viruses Ameliorate Gut Inflammation via Toll-like Receptor 3 and Toll-like Receptor 7-Mediated Interferon-β Production. Immunity **44**:889–900.

98. **Stecher B**, **Robbiani R**, **Walker AW**, **Westendorf AM**, **Barthel M**, **Kremer M**, **Chaffron S**, **Macpherson AJ**, **Buer J**, **Parkhill J**, **Dougan G**, **Von Mering C**, **Hardt WD**. 2007. Salmonella enterica serovar typhimurium exploits inflammation to compete with the intestinal microbiota. PLoS Biol.

99. **Lupp C**, **Robertson ML**, **Wickham ME**, **Sekirov I**, **Champion OL**, **Gaynor EC**, **Finlay BB**. 2007. Host-Mediated Inflammation Disrupts the Intestinal Microbiota and Promotes the Overgrowth of Enterobacteriaceae. Cell Host Microbe **2**:119–129.

100. **Schwille-Kiuntke J**, **Enck P**, **Zendler C**, **Krieg M**, **Polster A V.**, **Klosterhalfen S**, **Autenrieth IB**, **Zipfel S**, **Frick JS**. 2011. Postinfectious irritable bowel syndrome: Follow-up of a patient cohort of confirmed cases of bacterial infection with Salmonella or Campylobacter. Neurogastroenterol Motil.

101. **Thabane M**, **Simunovic M**, **Akhtar-Danesh N**, **Garg AX**, **Clark WF**, **Collins SM**, **Salvadori M**, **Marshall JK**. 2010. An outbreak of acute bacterial gastroenteritis is associated with an increased incidence of irritable bowel syndrome in children. Am J Gastroenterol.

102. **Rodríguez LAG**, **Ruigómez A**, **Panés J**. 2006. Acute Gastroenteritis Is Followed by an Increased Risk of Inflammatory Bowel Disease. Gastroenterology **130**:1588–1594.

103. **Kamdar K**, **Khakpour S**, **Chen J**, **Leone V**, **Brulc J**, **Mangatu T**, **Antonopoulos DA**, **Chang EB**, **Kahn SA**, **Kirschner BS**, **Young G**, **DePaolo RW**. 2016. Genetic and Metabolic Signals during Acute Enteric Bacterial Infection Alter the Microbiota and Drive Progression to Chronic Inflammatory Disease. Cell Host Microbe **19**:21–31.

104. **Small CL**, **Xing L**, **McPhee JB**, **Law HT**, **Coombes BK**. 2016. Acute Infectious Gastroenteritis Potentiates a Crohn's Disease Pathobiont to Fuel Ongoing Inflammation in the Post-Infectious Period. PLoS Pathog **12**:1–20.

CHAPTER 3

DYNAMIC CHANGES IN THE VIROME AND BACTERIOME IN PATIENTS FOLLOWING RECOVERY FROM ACUTE BACTERIAL DIARRHEAL ILLNESS

ABSTRACT

Acute bacterial gastroenteritis is a significant disease burden worldwide and affects all individuals of all ages. Gastroenteritis is primarily an acute, self-limiting infection, but it can be an initiating event for the onset of more chronic diseases like inflammatory bowel disease (IBD) or irritable bowel syndrome (IBS). Previous gastroenteritis studies identified increased Proteobacteria abundance in patients with active disease, specifically *Escherichia.* Few studies, however, have examined changes in the microbiome following recovery from an illness, and even fewer have evaluated the virome or the populations of viruses present in the microbiome. Herein, we have compared the composition of and changes within the microbiome among 79 patients (cases) with acute bacterial gastroenteritis to those from a subset (n=63) of the same cases post-recovery. Our findings further confirm an increased abundance of Proteobacteria in cases. Also, patients with similar microbiome profiles clustered together, and patients with microbial communities belonging to Cluster 2 were significantly more likely to have severe disease and more extensive microbiome alterations during infection compared to other cases. Three bacterial populations, *Alistipes*, *Sutterella*, *Odoribacter*, were lower in abundance in follow-ups compared to controls (Chapter 2), suggesting that these microbes may fail to recover following severe enteric infections. These microbes could be investigated as novel probiotics.

INTRODUCTION

The estimated incidence of acute gastroenteritis in the United States ranges from 179 million (1) to 375 million cases (2), with many cases unreported. Gastroenteritis can have two types of effects on human health. The first effect is an immediate acute illness, which involves pathogen infection followed by an expansion of Enterobacteriaceae populations and subsequent resolution (3). The second potential effect is a chronic, insidious inflammatory state that predisposes patients to post-infectious irritable bowel syndrome (IBS), which can result in symptoms for up to a decade following onset (4, 5) or inflammatory bowel disease (IBD). Individuals are 2.4 times more likely to develop IBD in the year following a case of infectious gastroenteritis (6).

Previous research has identified increased abundance of Proteobacteria in patients with acute bacterial gastroenteritis along with a decreased abundance of Firmicutes and Bacteroidetes (7). Another study identified the *Escherichia-Shigella* enterotype (8), which was defined as intestinal communities from gastroenteritis patients with an increased abundance of *Escherichia*, and was correlated with an over-abundance of *Veillonella* and *Staphylococcus* (8). Prior microbiome studies related to gastroenteritis have focused mainly on characterizing the composition of the bacterial populations during the infection. Few studies, however, have examined how the microbiome recovers following perturbations caused by infection with different bacterial pathogens. Consequently, we sought to determine how the viral and bacterial populations of the microbiome change during and after enteric infections. Classifying these changes will identify options to restore beneficial microbial communities that promote intestinal health.

MATERIALS AND METHODS

Sample selection and sequencing

Stool samples were obtained via an active surveillance system in coordination with the Michigan Department of Health and Human Services (MDHHS) and four Michigan hospitals (9). In total, 142 stool samples were utilized in this study; 79 samples were from patients with acute enteric infections (Chapter 2) and 63 from a subset of patients 1 to 26 weeks following recovery. Samples were transported in Cary Blair media, homogenized, centrifuged, and stored at -80 °C in triplicate. The QIAmp DNA Stool Mini Kit (QIAGEN; Valencia, CA) was used to extract DNA. Clinical details, demographics, and exposures were obtained for each patient using the Michigan Disease Surveillance System (MDSS). After recovery, cases were given a questionnaire regarding clinical symptoms and exposures. The Institutional Review Boards at Michigan State University (MSU; IRB #10-736SM), MDHHS (842-PHALAB), and the four participating hospitals approved all protocols.

A Perkin Elmer Sciclone NGS robot following the manufacturer's recommendations was used to prepare libraries for sequencing using an Illumina TruSeq Nano DNA Library Preparation Kit. Samples were added in duplicate for each sequencing run on from one of four equimolar library pools. Quality control of the libraries was done with qPCR and DNA was quantified with a Qubit dsDNA HS (Thermo Fisher Scientific, Waltham, MA, USA) and Caliper LabChipGX HS DNA (Caliper Life Sciences, Hopkinton, MA, USA). The library for Run 1 was sequenced in a 2x150 bp paired-end format after being loaded in two lanes of an Illumina HiSeq 2500 Rapid Run flow cell (v1) using Rapid SBS reagents. The libraries for Runs 2, 3, 4 were sequenced in 2x250 bp paired-end format after being loaded onto two lanes of an Illumina HiSeq 2500 Rapid Run flow cell (v2). The Illumina Real-Time Analysis (v1.18.61) was used for base-calling. The Illumina

Bcl2Fastq (v1.8.4) demultiplexed the output and converted it to fastq format. As previously described in Chapter 2. No significant differences were observed in the microbiome composition or sample clustering by sequencing runs.

Power analysis

The pwr package (10) in R (11), utilizes the power equations developed by Cohen (12). All power calculations made standard assumptions ($p = 0.05$, effect size = 0.5, power = 0.8) for each statistical test in the study (Chi-square, analysis of variance, correlation, regression). Power curves were generated to show the relationship between sample size and effect size (Figure 3.1). The power curves demonstrate that a minimum of 88 samples is needed to have enough power (80%) to detect difference between two study groups (cases and follow-ups). The study has 142 samples (power=94%), thus we have adequate power to detect differences between the two study groups (cases and follow-ups).

Sequence processing and metagenomics

Processing and annotation Trimmomatic (13) was used to remove sequencing adaptors, and low-quality reads. FastQC (14) generated a report regarding read quality, adaptor contamination and GC skew for sequenced reads. The following pipeline is based on on Norman *et al.* (15) and KBase (16). In brief, quality-controlled (per base sequence quality $\geq$ 30) reads were aligned to a human RefSeq database (GRCh38_1118, downloaded November 2018) available at the National Center for Biotechnology Information (NCBI). using Bowtie 2 (17) and SAMtools (18). Quality-controlled reads were annotated using Kaiju (19). Kaiju aligns each read to the non-redundant protein database (20) of viruses, bacteria, and fungi (nr_euk, downloaded January 2019) at NCBI. The results are then summarized to create a microbiome profile across different taxonomical ranks. On average 86% of reads were annotated at the phylum level, 62% at the genus level, and 26% at

the species level (Figure 3.2). Custom python (21) scripts were employed that mimic Kaiju's *kaiju2table* function and split the output into viral and non-viral annotations. Subsequent analysis was performed at the following taxonomical levels (Phylum, Class, Order, Family, Genus, Species), as done previously (22).

Reads were assembled with Metaspades (23). Reads were mapped to assemblies to assess assembly quality and on average, 10.9% of reads in cases, and 9.8% of reads in follow-ups did not map to the assemblies, which was not statistically significant (Mann Whitney U test $p = 0.1146$). Additional analysis was conducted the assessed statistical trends between sequencing depth, coverage, and alpha diversity metrics and no trends were noted between these variables ($R^2 < 0.7$, spearman $p > 0.05$).

The maximum number of reads (paired-forward) sequenced in a sample was 7,427,518 (3.7 Giga base pairs [Gbp]) out of all samples (n=142) and the average sequencing per sample was 2,967,423 reads (1.2 Gbp). No significant differences were noted between the sequencing depth for cases and follow-ups (Mann Whitney U test $p = 0.3492$). Rarefaction (24) measures species richness, and rarefaction curves (25, 26) of genera data were created with the *rarefy* and *speccacum* functions from the vegan package (27) in R. Sequencing depth was sufficient for both cases and follow-ups since both species' accumulation curves (random sampling, Figure 3.3A) and rarefaction, Figure 3.3B) achieved plateau. Nonparielle3 (28) calculated coverage for each sample. The average coverage for all samples (n=142) is 80% based on Nonparielle3. The genus-level classification was used for analysis as was done in Chapter 2. Scripts are available at github.com/BrianNo.

Cluster analysis Microbial taxa that were not present in at least 1% of samples were removed to reduce the false-positive rate of genera significance as recommended (22, 29). Zero counts in the

taxonomy table were replaced using multiplicative simple replacement with the zCompositions package (30) in R, which attempts to estimate a small value for zero based on the values in the table. MixOmics (29) calculated the relative abundance as a percent of the total annotated viral and bacterial populations, based on (31). A compositional-data-analysis approach was undertaken (32) and a center-log-ratio transformation was performed on the relative abundance. Hierarchical clustering of transformed microbiome profiles was performed using Ward's linkage and Aitchison distance (33) between samples.

The *adonis* function from vegan package in R performs non-paremetric MANOVA (NP_MANOVA) and determined if microbiome profiles were different based on cluster and case status. P-values for multiple hypothesis testing were adjusted using a Bonferroni correction with the *p.adjust* function. The *betadisper* function from the vegan package assessed group heterogeneity. Analysis of composition of microbiomes (ANCOM) (34) determined the taxa that were differentially abundant between groups (clusters, case status). A network analysis was performed following the Spieceasi pipeline (35) and correlations calculated between different taxa with SparrC (36) correlated different taxa with one another to create a taxonomical network. The alpha diversity (Shannon index), Richness (total number), and Evenness (distribution) at the genus level were calculated based on the read count of each taxonomical assignment using the vegan package in R.

Data analysis

Microsoft Excel and Access were used to manage demographic and epidemiological data All statistical analysis was performed using R and EpiInfo (37). Univariate analysis was performed using chi-square and Fisher's exact tests (for when counts ≤ 5) to identify associations between the exposure (independent) and outcome (dependent) variables; p-value < 0.05 are considered

significant. Exposure variables include epidemiologic, exposure, and demographic data. Outcome variables include case status, cluster membership, and disease severity.

Univariate variables that had strong associations ($p \leq 2.0$) with outcomes of interest were included in the multivariate logistic regression model. A forward and backward selection was used to build the model. Variables such as age, sex, race, and infection type, were included in the model. Factors were then added or removed if significant changes occurred in the model ($p < 0.05$). Each factor was assessed for collinearity. The statistical significance of each coefficient in the model was assessed with the Wald test. The Hosmer-Lemeshow test (38) assessed the goodness of fit.

RESULTS

Case and follow-ups had different viral and bacterial read counts

In total, 621,384,080 (189.2 Gbp) paired forward reads were sequenced across all 142 samples, yielding 3,046,000 or 1.4 Gbp paired-forward reads per sample. Cases and follow-ups achieved average sequencing depths of 3,041,142 reads (1.4 Gbp) and 2,874,981 reads (1.4 Gbp), respectively, with no difference between study groups (Mann Whitney U test p = 0.3492). The average coverage, as determined by Nonpareil3 (28), was 80% across all samples. Although cases had lower coverage (77%) than follow-ups (83%), the difference was not statistically significant (Mann Whitney U test p = 0.349). On average, across all samples, 12.9% of reads fell below quality filtering parameters. More reads were removed from cases (13.3%) compared to follow-ups (12.4%) though this difference was also not significant (Mann Whitney U test p = 0.2195).

On average, 6% of all quality-controlled reads were annotated as human derived. The abundance of human DNA differed by sample type; cases comprised 15.2% human reads compared to only 0.1% in follow-ups, which was a statistically significant difference (Mann Whitney U test p = 9.343e-12). Kaiju annotated 61.5% of reads to the Genus level that passed quality control (i.e., trimming and human read removal steps) across all samples. Follow-up samples achieved a higher annotation frequency (64.2%) compared to case samples (59.3%), the difference in annotation frequency was significant (Mann Whitney U test p = 0.01802). On average, 61% of reads were annotated to bacteria across all samples at the Genus level, and 0.45% of reads were assigned to viruses. Cases had a significantly lower proportion of reads assigned to bacteria (58.7%) compared to the follow-ups (64%; Mann Whitney U test p = 0.008385). Cases also had an increased proportion of viruses (0.7%) compared to follow-ups (0.2%), though this difference was not statistically significant (Mann Whitney U test p = 0.1449) (Table 3.1.). Case

communities were not significantly different in diversity using the Shannon index (Mann Whitney U test p = 0.4139, Figure 3.4A) but were significantly different in richness (Mann Whitney U test p = 1.273e-07, Figure 3.4B) when compared to follow-ups at the genus level. Evenness was not significantly different between cases and follow-up (Mann Whitney U test p = 0.8631, Figure 3.4C).

Description of Cohort

Between January 2011 and December 2015, stool samples were recovered from 79 patients with enteric infections and a follow-up sample was obtained from 63 of these cases after recovery from illness between 1 to 26 weeks post-infection. Among the cases included in this analysis, 48.1% (n=38) were males while 51.9% (n=41) were female (Table 3.2.). Most of these cases were between 19 and 64 years of age (n=33, 41.8%) followed by 0 to 9 years (n=21, 26.6%). Cases reportedly resided in Oakland (n=20; 17.2%), Wayne (n=20; 17.2%), Ingham (n=15, 12.9%), and Eaton (n=13, 11.2%) counties; 48.7% (n=38) of these cases were from an urban residence compared to 51.3% (n=40) from a rural residence. Among the 79 cases, *Salmonella* spp. was the most common infection (n=35, 44.3%) followed by *C. jejuni* (n=29, 36.7%), *Shigella* spp. (n=10, 17.7%), and STEC (n=5, 6.3%). The most common symptoms were body aches (n=73, 94.8%), fever (n=47, 69.1%) and vomiting (n=47, 69.1%). In all, 37.7% (n=29) of the cases were hospitalized, with 53.6% (n=15) of the cases requiring hospitalization for more than two days (Table 3.2.).

Among the cases who submitted follow-up samples, 44.4% (n=28) were male and 55.6% (n=35) were female (Table 3.2.). The highest frequency of samples was collected from the 19-64 age group (n=26, 41.3%); the second highest was in the 0-9 age group (n=17, 27%). The highest counties samples were collected from were Ingham (11,17.7%) and Washtenaw (10, 16.1%).

46.8% (n=29) of follow-up samples were from an urban residence and 53.2% (n=33) from a rural area (Table 3.2.). Each follow-up sample was cultured and confirmed to be negative for the pathogen associated with the original infection. These follow-up samples were submitted for patients originally infected with *Salmonella* spp. (n=28, 44.4%), *C. jejuni* (n=25, 39.7%), *Shigella* spp. (n=7, 11.1%) and STEC (n=3, 4.8%).

Microbiome composition varies between patients and follow-ups

In total, 473 (449 bacterial, 24 viral) Families were identified among all 142 samples. At the genus level, there were 2,659 genera identified (2,482 bacteria and 177 viruses). Five virus families including Myoviridae, Poxviridae, Microviridae, and Siphoviridae were found to be the most differentially abundant between cases and follow-ups (Figure 3.5A). Myoviridae and Poxviridae were more abundant in cases comprising 26% and 9% of viral reads on average compared to follow-ups, which comprised 19% and 1% of reads, respectively. Poxviridae was significantly higher in cases compared to follow-ups (Mann Whitney U test p = 5.404e-09). By contrast, Microviridae and Siphoviridae were more abundant in follow-ups comprising 17% and 41%, respectively, compared to cases on average. Microviridae was significantly lower in cases (6%) compared to follow-ups (17%, Mann Whitney U test p = 0.0008933).

Bacterial profiles were distinct between the case and follow-up samples as well (Figure 3.5B). Examination of the top 10 differentially abundant bacterial families showed that Enterobacteriaceae were significantly more abundant in cases with bacterial reads accounting for 34.4% of the total bacterial reads on average. This level was significantly different from the average level (2.9%) observed for follow-ups (Mann Whitney U test p < 2.2e-16). Bacteroidaceae, Ruminococcaceae, Rikenellaceae, and Prevotellacea were also significantly more abundant in the

recovered samples, on average, than in the case samples, accounting for 48%, 9%, 7%, and 4%, respectively (Mann Whitney U test p =0.008184, 2.971e-06, 0.01456, 0.003112).

Hierarchical clustering generates four distinct clusters

Cluster 1 (n=27) consists of 40.7% males (n=11) and 58.3% females (n=16). Cluster 2 (n=33) has 54.5% males (n=18) and 45.5% (n=15) females. Cluster 3 (n=22) has 59.1% males (n=13) and 40.9% (n=9) females. Cluster 4 (n=60) consists of 40% males (n=24) and 60% (n=36) females, (Table 3.3.). Additionally, the 19-64 age group is the most common age group across all 4 clusters; Cluster 1 (n=12, 44.4%), Cluster 2 (n=13, 39.4%), Cluster 3 (n=8, 36.4%), Cluster 4 (n=26, 43.3%). The second most common age group is 0-9 across all 4 clusters; Cluster 1 (n=8, 29.6%), Cluster 2 (n=8, 24.2%), Cluster 3 (n=7, 31.8%), Cluster 4 (n=15, 25%), (Table 3.3.) Clusters vary in accordance to health state. Cluster 1 (n=27) consists of 93% cases (n=25) and 7% follow-ups (n=2). Cluster 2 (n=3) is 100% cases (n=33). Together (clusters 1 and 2) accounted for 73% (n=58) of the cases (n=79). Cluster 3 (n=22) is 31.8% cases (n=7) and 68.2% follow-up (n=15). Cluster 4 (n=60) consists of 23.3% cases (n=14) and 76.7% (n=46) follow-ups. Together (clusters 3 and 4) account for 97% (n=61) of the follow-ups (n=63), (Table 3.2). Case hospitalization rates were varied across clusters. Cluster 1 (n=6, 24%), Cluster 2 (n=16, 48.5 %), Cluster 3 (n=2, 28.6%), Cluster 4 (n=5, 38.5%). Reported symptoms (only available for cases in each cluster) were varied across clusters. This collective data is similar to the results with cases and controls (Chapter 2), suggesting further that the follow-up state is similar to controls. The PCA demonstrates that Cluster 3 and Cluster 4 are mainly localized on the right side and represent the majority of follow-up samples, (Figure 3.6A). Cluster 2 is located most distally on the left side of the PCA with some overlap into Cluster 1 which is localized medially, (Figure 3.6A). Cluster 3 (purple) is the most mixed cluster (31.8% case and 68.2% follow-up) and is the most

heterogeneous (Figure 3.6B), followed by the most distant cluster, cluster 2 (100% cases). Clusters are considered distinct (permanova $p < 0.001$).

Shannon index for diversity did not vary across clusters (Kruskal Wallis test $p = 0.1787$). However, trends were noted, Cluster 4, which had the most follow-ups, had the highest diversity. Cluster 2 (100% cases) had the lowest diversity (Figure 3.7A). Richness was significantly different across clusters (Kruskal Wallis test $p = 2.573e\text{-}13$). Case-dominated clusters (Clusters 1 and Cluster 2) had lower richness compared to the recovered, follow-up-associated clusters (Clusters 3 and Cluster 4) (Figure 3.7B). Cluster 2 (100% cases) had the lowest richness of all clusters. Cluster 4 had the highest richness. Evenness was not significantly different across clusters (Kruskal Wallis test $p = 0.7993$ (Figure 3.7C)). Collectively, this data shows that clusters represent the collective differences observed in cases based on cluster membership. We then sought to identify disease associations with each case-dominated cluster.

Gastroenteritis symptoms are associated with microbiomes from cases

Univariate analysis with chi-square was identified disease associations with each cluster that was dominated by cases (Cluster 1 or Cluster 2). Each clinical characteristic (i.e., symptoms, hospitalization status) was considered an exposure (independent variable) and the cluster (Cluster 1 or Cluster 2) to which the sample belong was the outcome (dependent variable). Cluster 1 is localized medially on the PCA, (Figure 3.6A). We hypothesize less severe illness (no bloody diarrhea, chills, fever) and more non-specific (abdominal pain, nausea, fatigue). Cluster 1 was found to be associated with body aches (OR: 7, CI (95%): 2.4, 20.8), (Table 3.4.). Cluster 2 was the most distant cluster on the PCA; we hypothesized that because of this distance (representing dysbiosis), Cluster 2 would have associations with more severe disease (bloody diarrhea, fever, chills, vomiting). Cluster 2 was associated with vomiting (OR: 5.7, CI (95%): 2.1, 15.8), and was

trending (p-value < 0.2) with bloody diarrhea (OR: 2.1, CI (95%): 0.8, 5.5), and headache (OR: 2.3, CI (95%): 0.8, 6.6). Additionally, hospitalization was not associated with Cluster 1 (OR: 1.1, CI (95%): 0.4, 3.2), though patients with Cluster 2 communities were trending (p < 0.2) to more likely be hospitalized (OR: 2.3, CI (95%): 0.9, 5.9), (Table 3.5.). No associations were found with sex, gender, or race for either Cluster 1 or Cluster 2. Each cluster had different disease associations, so we then decided to determine the organisms that were distinct in each from the rest of the study.

Specific viruses and bacteria are associated with either cluster

Cluster 1 (green) has a composition more like that of clusters 3 and 4 with minor alterations, (Figure 3.6A). However, there are distinct differences in Cluster 1, which also make it similarly to Cluster 2. Cluster 2 (orange) has the most distinct microbiome compared to the other three clusters. There are increases and decreases in several taxa, (Figure 3.8). The increased abundance of these taxa could represent a bloom in these microorganisms during infection. ANCOM was utilized to determine the microbial composition that is unique to Cluster 1 and Cluster 2. It was hypothesized that Cluster 1 and Cluster 2 would share some taxa in common, but Cluster 2 will have a very different profile, which will include Enterobacteriaceae, bacteriophages related to Enterobacteriaceae and eukaryotic viruses.

ANCOM identified seven differentially abundant taxa in Cluster 1, 92 for Cluster 2, and seven of the taxa were shared between both clusters. All the taxa that were determined to be important for Cluster 1 are also found in Cluster 2. The microbiome that is shared (n=7 genera) between Cluster 1 and Cluster 2 is dominated by Proteobacteria (n=6 genera, 86%) and includes genera representing the common enteric pathogen, *Salmonella*, as well as other pathogens (*Enterobacter*, *Citrobacter*, *Hemophilus*) and Firmicutes (n=1 genus, 14%, *Raoultella*), (Table 3.6.). Cluster 2 is defined by 85 additional genera distinct from the common microbiome between

Cluster 1 and Cluster 2. Viruses comprise 34.1% of this difference (n=29 genera). 89.6% of the viral taxa identified unique to Cluster 2 are Caudovirales (n=26 genera) and include; Podoviridae (n=6 genera), Siphoviridae (n=12 genera), Myoviridae (n=7), Helleviridae (n=1 genus), and eukaryotic virus (n=3), (Table 3.6.). Bacteria genera that are differentially abundant and unique to Cluster 2 (n=56 genera) consist of 64.2% Firmicutes (n=36 genera), 5.4% Bacteroidetes (n=3 genera), and 25% Proteobacteria (n=14 genera), and 5.4% Others (n=3 genera), (Table 3.6.). Network analysis demonstrates that many of the genera identified differential for Cluster 2 is strongly correlated, (Figure 3.9). The Second highest differentially abundant (green) were positively correlated with one another as well Tissierellales (*Tissierella*), Lactobacillales (*Lactobacillus*), surprisingly Bacteriodales were not directly correlated. Clostridales were most negatively correlated (red edges) with Enterobacterales (red) and other pathogenic bacteria, including Bascillales (*Staphylococcus*) Lactobascillales (*Enterococcus* and *Streptococcus*), and Psuedomonindales (*Acinetobacter*, *Psudomondonas*) as well Caudovirales that infect Enterobacterales, Lactobacillaes, Bacillales, Pseudomondales. Eukaryotic viruses (*Orthopoxvirus*, *Cytomegalovirus*, and *Mastadenovirus*) were also negatively correlated with Clostridales.

Enterobacterales was the other most highly connected part of the network and was positively correlated with other genera commonly representing pathogens, including Psuedomonindales (*Acinetobacter*, *Psudomondonas*), Lactobascillales (*Enterococcus*, and *Streptococcus*), and Bascillales (*Staphylococcus*). Enterobacterales was positively correlated with some Enterobacterales viruses like *P2virus* and *Nona33virus* and eukaryotic viruses *Orthopoxvirus*, *Cytomegalovirus*, and *Mastadenovirus*.

A univariate analysis was performed in order to identify taxa that are in higher abundance (i.e., blooming) for Cluster 2 and subsequently change post-recovery. ANCOM analysis was

utilized in the selection of taxa, and factors were generated that stated if a sample was above or below the normalized average for a given taxon. Taxa were the exposure (independent variable) and the presence of a sample in Cluster 2 was the outcome (dependent variable). Cluster 2 was found to be associated with the following virus genera that were above the study average; *Orthopoxvirus* (OR: 15.2, CI (95%): 5.9, 39.5), representing 66.7% (n=20) of the total above average (n=30), and *Cytomegalovirus* (OR: 11, CI (95%) 2.4,69.1), common enteric bacteriophage, *Nona33virus* (OR: 13.2, CI (95%): 3.9, 52.5), representing 72.2% (n=13) of the total above average in study (n=18), *Lambdavirus* (OR: 8.8, CI (95%): 2.5, 36.3), *P22virus* (OR: 4.6, CI (95%): 1.1, 20.4), and *P2virus* (OR: 4.6, CI (95%): 1.4, 14.9), *Kayvirus* (OR: 2.9, CI (95%): 1, 7.8), *Seuravirus* (OR: 5.7, CI (95%): 1.3, 29.7), and *Np1virus* (OR: 5.5, CI (95%): 1.7, 17.3).

Associations were observed for bacterial genera that were above average in the study including members comprising common enteric pathogens such as *Salmonella* (OR: 3.6, CI (95%): 1.2, 10.1), *Escherichia* (OR: 14.1, CI (95%): 5.6, 35.8), and *Shigella* (OR: 22.2, CI (95%): 8.2, 60.1). Other genera comprising pathogenic bacteria including *Enterobacteria* (OR: 30.6, CI (95%): 9.3, 123.5), which represented 80% (n=20) of the total above average (n=25), *Pseudomonas* (OR: 6.2, CI (95%): 1.1, 42.4), *Staphylococcus* (OR: 14.9, CI (95%): 4.4, 59.3), *Haemophilus* (OR: 4.3, CI (95%): 1.6, 11.5), *Acinetobacter* (OR: 37.9, CI (95%): 7.8, 369.5) representing 87.5% (n=14) of the total above average (n=16), *Streptococcus* (OR: 4.3, CI (95%): 1.6, 11.5), *Klebsiella* (OR: 4.8, CI (95%): 1.8, 13.2), *Vibrio* (OR: 11.2, CI (95%): 3.8, 32.8), *Enterococcus* (OR: 18.7, CI (95%): 5.2, 86.5), other bacteria including; *Citrobacter* (OR: 3.7, CI (95%): 1.4, 10.1), *Pantoea* (OR: 204, CI (95%): 29.7, 8367.8) which represents 69.6% (n=32) of the total above average (n=46), *Raoultella* (OR: 6.2, CI (95%): 1.1, 42.4), *Peptostreptococcus* (OR: 14, CI (95%): 2.5, 146.1), *Hafnia* (OR: 10.6, CI (95%): 0.8, 570), and *Serratia* (OR: 31.9, CI (95%): 8, 188.9), (Table

3.7.). Based on these different taxa associations with Cluster 2 and network analysis (Figure 3.9), we sought to build a model that incorporated these associations to define the most important members that could predict Cluster 2 status.

Logistic Regression Modeling for predicting Cluster 2 status

Pantoea was selected as the base-model because it had the greatest odds-ratio, the highest number of observations above average across samples, and was an integral member of the network analysis. Logistic regression was performed, additionally, with gastroenteritis-causing organisms (*Salmonella*, *Shigella*, *Escherichia*), of which *Shigella* was determined to have the most significant contribution to the model. Opportunistic pathogens were assessed and included (*Enterobacter*, *Serratia*, *Enterococcus*); Enterobacter was selected for final *model* due to improvements that it provided. Bacteriophages that directly infect Proteobacteria *(*P22viurs, *Nona33virus*, *Lambdavirus*) and Eukaryotic viruses (*Orthopoxvirus*, *Cytomegalovirus*) were tested and did not substantially improve the model. Wald's test was used to incorporate significant variables. The Hosmer-Lemeshow goodness-of-fit test and the AIC were both evaluated to determine if the model was being overfitted. Wald's test was used to incorporate significant variables. The final model (model 9) consists of *Shigella*, *Enterobacter*, *Pantoea* in defining Cluster 2 status, (Table 3.7.).

Matched cohort further confirms previous findings

We then performed a matched case follow-up cohort analysis to investigate the differences and examine dysbiosis longitudinally within the same individual. In total, there were 62 matched case and follow-up samples that were matched (n=124). No statistical differences in sequencing quality were found between this matched cohort (n=124) and the cohort used earlier (n=142) nor within the matched cohort. The total number of bacterial and viral Families (n=473) and Genera (n=2,659) were the same between both matched and unmatched cohorts. The Shannon index did

not vary between the case and follow-up states within the matched cohort (Wilcoxon Signed-rank test p = 0.3644); however, genera richness was significantly increased in follow-ups compared to cases (Wilcoxon Signed-rank test p =1.96e-06). Evenness was not statistically different between cases and follow-ups (Wilcoxon Signed-rank test p = 0.9144).

Differential abundance analysis identified with ANCOM identified that 36 genera were differentially abundant between the matched cases and follow-ups. In total five genera were viruses (*P22virus*, *P2virus*, *Nona33virus*, *Orthopoxvirus*, *P1virus*) and 31 genera were bacteria. Among these bacterial genera, 16 were Proteobacteria (*Salmonella*, *Escherichia*, *Shigella*, *Klebsiella*, *Campylobacter*, *Citrobacter*, *Haemophilus*, *Vibrio*, *Pantoea*, *Acinetobacter*, *Pseudomonas*, *Atlantibacter*, *Proteus*, *Hafnia*, *Providencia*, *Morganella*), 14 were Firmicutes (*Roseburia*, *Veillonella*, *Flavonifractor*, *Subdoligranulum*, *Anaerotruncus*, *Pseudoflavonifractor*, *Staphylococcus*, *Oscillibacter*, *Intestinibacillus*, *Intestinimonas*, *Anaeromassilibacillus*, *Lawsonibacter*, *Neglecta*), and one was Actinobacteria (*Rothia*).

Examination of the top differentially abundant viruses (Figure 3.10A) demonstrated that *Orthopoxvirus* (Wilcoxon signed-rank test p = 0.0002967) and *Nona33virus* (Wilcoxon signed-rank test p = 0.003692) were significantly increased in abundance in cases compared to the follow-up samples. *Felixo1virus* (Wilcoxon signed-rank test p = 0.005037) and *Seuratvirus* (Wilcoxon signed-rank test p = 0.001004) were increased in abundance in follow-ups compared to cases. Examination of the bacterial changes (Figure 3.10B) demonstrates that Proteobacteria such as *Escherichia* (Wilcoxon signed-rank test p = 1.878e-08), *Shigella* (Wilcoxon signed-rank test p = 1.201e-07), and *Salmonella* (Wilcoxon signed-rank test p = 1.86e-11) were significantly more abundant in cases than follow-ups. The follow-up samples, however, had significantly more abundant *Bacteroides* (Wilcoxon signed-rank test p = 0.005151) and Firmicutes such as *Roseburia*

(Wilcoxon signed-rank test p = 1.801e-06), *Alistipes* (Wilcoxon signed-rank test p = 0.002572), *Akkermansia* (Wilcoxon signed-rank test p = 0.001229), and *Ruminococcus* (Wilcoxon signed-rank test p = 0.0001088) relative to the cases.

DISCUSSION

The resident microbes are continually changing, and subsequently the microbiome is as well. Studies have found that an insult can perturb the microbiome and drastically change its composition. Studies that have followed the microbiome through a time course have linked changes to diet (39), antibiotic-use (40), pregnancy (41), prediabetes (42), and IBD (43). IBD patients (43) were found to have microbiome shifts that consistently show a lower abundance of *Faecalbacterium, Subdogligranulum,* and *Roseburia* which are considered important for a healthy microbiome. In the same study (43), *E. coli*, *Haemophilus parainfluenzae,* and *Klebsiella pneumoniae* were all increased in abundance, which are negatively correlated with health. Here we have observed similar findings as Roseburia (0%) was not elevated in communities belonging to Cluster 2, though *Haemophilus* (50%) and *Klebsiella* (52%) were elevated in Cluster 2 which represents microbes potentially blooming during the acute infection.

There have been few studies that have examined acute bacterial gastroenteritis before and after an infection. Previously, a 16S rRNA sequencing analysis of 310 samples collected through our ERIN study showed expansions in Proteobacteria, specifically *Escherichia*, during acute bacterial gastroenteritis regardless of the agent causing each bacterial infection (7). Moreover, the abundance of Proteobacteria was found to decrease post-recovery, which was similar to the levels observed for uninfected, healthy individuals (7). Here we utilized metagenomics, which offers greater resolution of resident microbial communities, and observed similar trends. Specifically, we found an increased abundance of *Escherichia* among patient samples compared to the follow-up samples submitted by the patients following recovery from infection. No significant difference in the abundance of *Escherichia* was observed in these follow-up samples compared to those from healthy individuals (Chapter 2), suggesting that decreases in *Escherichia* may be important for

recovery. Additional bacterial populations, namely *Alistipes*, *Sutterella*, *Odoribacter*, were also found to be lower in abundance among the follow-up samples compared to controls (data not shown, Chapter 2). *Alistipes* abundance has been correlated positively with health (44), while *Odoribacter* represent a group of butyrate producers known to regulate inflammation (45). Similarly, *Sutterella* is a common commensals that could aid in immune regulation and function (46). Given that these microbial populations were lower in follow-up versus healthy samples, suggesting that these bacteria might be lost or failed to recover following infection. Therefore, these microbes should be investigated in future studies as potential targets for probiotic therapy to promote recovery from gastroenteritis.

Fewer studies have directly studied changes in the virome in gastroenteritis. Previous studies have found that the virome has a high degree of variation between individuals, and the virome is conserved over time (47–49), suggesting the possibility of a core virome. It has been shown that the virome even stays constant after a fecal microbiota transplant (50). Studies have found increased abundance of Caudovirales, tailed bacteriophage, with chronic inflammatory disease like IBD (51), and increased abundance of adenovirus in HIV positive patients was associated with lower CD4 counts (52); however, neither of these studies followed the virome longitudinally. We sought to classify the alterations in the microbiome among patients with acute bacterial gastroenteritis during a disease state and a recovery state. The use of enterotypes (53) allows for clustering of samples based on common microbial compositions. Four clusters were identified; two case-associated and two follow-up-associated (recovery). Cluster 1 was associated with minor illness and associated more with the follow-up clusters; less severe disease has been noted in patients that had microbial profiles more similar to uninfected controls (54). Cluster 2 was associated with more severe illness and was more distant on the PCA with observable

differences notable on the heatmap. The taxa that were differentially abundant for Cluster 1 were also found in Cluster 2 and included the Proteobacteria (*Salmonella, Citrobacter, Haemophilus, Enterobacter, Kluverya, Pantoea*,) and Firmicutes (*Raoultella*). Salmonella has been shown to outcompete host microbes by exploiting inflammation (55). Additionally, *Salmonella* and *Citrobacter* have both been shown to induce inflammatory states in mouse models that allowed for the expansion of *Escherichia* to maintain inflammation long after the initial microbe cleared (56). *Haemophilus* is pro-inflammatory (57) and commonly has been associated with hospitalization (58). *Haemophilus* has been associated with other illnesses including multiple sclerosis (59), rheumatoid arthritis (60), colorectal carcinoma (61), and gastroenteritis (8, 62). *Pantoea* was the most surprising finding as it can cause disease in humans (63), but little is known about pathogenesis or its role in the microbiome. Among all taxa identified, *Pantoea* had the highest association (OR: 204, 29.7 - 8376.8) with Cluster 2. There was no difference between *Pantoea* abundance in follow-up and control samples (data not shown), suggesting that *Pantoea* most likely represents an opportunistic pathogen that temporarily blooms in cases which may be lost when transitioning from the follow-up vs control states.

Viruses (29 in total) and bacteria (56 in total) dominated the additional taxa. The 56 bacteria can be associated with disease or a healthy state. The majority of taxa that were differentially abundant in Cluster 2 are pro-inflammatory and include *Escherichia* and *Shigella*, both commonly elevated in gastroenteritis (7, 8), *Staphylococcus* has been associated with abdominal pain (54), but abdominal pain was not associated with Cluster 2. *Acinetobacter* had increased abundance in Cluster 2, and evidence suggests that it causes differentiation of T cells *in vitro* and downregulates helper T cells (64), potentially changing the immune system response to the dysbiosis.

Enterococcus produces bacteriocins that have strong antimicrobial properties (65), which can impact the growth of other bacteria and are elevated in Cluster 2.

Bacteria commonly associated with good health were found to be decreased in Cluster 2 and include; *Subdogligranulum*, *Gemmiger* (43), and *Roseburia*, (66), all of which are butyrate producers that have been shown to decrease inflammation (67). Subdogligranulum is increased in abundance after supplementation with *Lactobacillus* (68, 69), a common probiotic. There are many changes to viral composition as well in Cluster 2, most directly infect Enterobacteriaceae.

Caudovirales was increased for both gastroenteritis patients and Cluster 2; similar findings were seen in a study on IBD, phage increased in abundance and diversity within IBD patients, while the bacterial population was conversely decreased (51) and blooms in bacteriophage have been tied to increases in host inflammation (71) and affect the bacterial population directly (72). Expansion in Caudovirales have also been noted in viromes of immunocompromised HIV-infected patients, who have altered pro-inflammatory microbiomes and expanded *Adenovirus* populations (52). It has been proposed that Caudovirales can potentially control blooms in bacteria populations (73) through several mechanisms including adhering to mucosa (74) and direct population control through prophage induction (75). Our findings support these models in that the presence of Enterobacteria-phage is present alongside increases in their host, *Enterobacteriaceae*. Further analysis, however, will be needed to identify if the bacteriophages are lytic or lysogenic.

Elevations in eukaryotic viruses such as *Mastadenovirus* and *Cytomegalovirus* have been observed in gestational diabetes dysbiosis (76), but their role remains ill-understood but was suggested that they could generate a pro-inflammatory environment. *Orthopoxvirus* was also found to be elevated in abundance in gestational diabetes (76). *Orthopoxvirus* produces molecules which can bind cytokines, chemokines, and interferon to lower the immune response (77, 78).

Mouse models infected with *Orthopoxvirus* have elicited distinct changes in the microbial profiles, which include decreases in Proteobacteria abundance compared to mock (79). Additional murine models have shown that eukaryotic viruses can alter the host immune system. Murine norovirus in germ-free mouse models restored the morphology of the intestinal tract through a signaling cascade without an immune response to the virus (80) this suggests that some eukaryotic viruses support bowel homeostasis and might be integral to its regeneration after being damaged. Additionally, inactivated rotavirus has been shown to reduce inflammation in the colon through the induction of anti-inflammatory cytokines acting on toll-like receptors (81).

It should be noted that studies have identified that Poxviridae can be a false positive (52). However, previous studies utilized a viral only databases with BLAST at a standard e-value (10^{-5}); given the smaller database size of the viral only databases a more significant e-value should be utilized in studies that utilize a viral only database since e-value is calculated based on the database size; tools have been developed which attempt to ameliorate this issue (82). Here we used the entire NCBI non-redundant database (Bacteria, Viral, Eukaryotic) with a kmer-based annotation approach. In this study, the identification of a sequence as *Orthopoxvirus* has to have a higher score compared to all non-viral signatures in the database. Additional analysis is needed into the presence of *Orthopoxvirus* to confirm that it is a false positive of something known or unknown or is indeed *Orthopoxvirus*. Additionally, the biological significance of this virus needs to be investigated.

The study here is limited due to sample size (n=142) and the timing of the follow-up samplings, which were inconsistent and varied between 1 and 26 weeks. Neveretheless, the study analyzed both the viral and bacterial signatures simultaneously in acute bacterial gastroenteritis in a patient and their recovered state. Our overall findings indicate that there might be two subtypes

of acute bacterial gastroenteritis with different microbial profiles and disease presentations. The more dysbiotic profile had a more substantial proportion of Caudovirales, which correlated negatively with decreases in healthy bacteria such as *Subdoglibgranulum* and *Gemmiger* and correlated positively with increases in inflammation-promoting bacteria like *Shigella* and *Escherichia*. The logistic regression model identifies *Shigella*, *Enterobacter*, and *Pantoea* as being able to identify Cluster 2 status. Interestingly Enterobacter and Pantoea were also found to be differentially abundant in Cluster 1 as well. *Enterobacter* and *Pantoea* could be critical changes in the microbiome of acute bacterial gastroenteritis.

APPENDIX

Table 3.1. Sequencing quality and coverage estimates for 142 metagenomes Results for the total sequencing (column 2), the quality control (columns 3-4), annotation results (column 5) and overall coverage (column 6).

Study ID	Reads Paired-forward total Count (Gbp)	Reads remaining after low-quality read removal Count (%)	Reads remaining after human read removal Count (%)	Reads annotated Total (%) Viral (%)	Nonpareil Coverage (%)
ER0106	2891668 (0.9)	1981379 (68.5)	1977478 (99.8)	66.39 (0)	0.96
ER0110	4360036 (1.3)	2977906 (68.3)	2977572 (100)	65.94 (0.04)	0.88
ER0162	3277708 (1)	2205167 (67.3)	2204701 (100)	70.19 (0)	0.86
ER0187	4234852 (1.3)	2964081 (70)	2963580 (100)	66.02 (0.01)	0.86
ER0214	4248380 (1.3)	2935060 (69.1)	2899595 (98.8)	69.58 (0.01)	0.85
ER0221	3387746 (1)	2242100 (66.2)	2241949 (100)	59.87 (0.01)	0.81
ER0227	985823 (0.3)	654468 (66.4)	654264 (100)	66.78 (0.01)	0.62
ER0235	1151295 (0.3)	748995 (65.1)	745840 (99.6)	53.1 (0.01)	0.58
ER0239	260491 (0.1)	145314 (55.8)	144650 (99.5)	63.57 (0.63)	0.3
ER0288	5243347 (0)	3333372 (63.6)	3328755 (99.9)	45.63 (2.56)	0.78
ER0302	5640052 (0)	3822002 (67.8)	3821846 (100)	74.35 (0)	0.93
ER0378	3457684 (1.6)	2334944 (67.5)	2334771 (100)	59.1 (0)	0.87
ER0437	591000 (1.7)	343633 (58.1)	343596 (100)	75.29 (0.02)	0.77
ER0689	2184150 (1)	1511874 (69.2)	1507591 (99.7)	72.24 (0.01)	0.82
ER0127	2704985 (1.2)	2188578 (80.9)	2186351 (99.9)	54.19 (0.06)	0.91
ER0137	2513753 (0.7)	2429928 (96.7)	2429913 (100)	63.04 (0.05)	0.88
ER0153	2131633 (1.4)	2047147 (96)	2046192 (100)	68.5 (0.02)	0.85
ER0269	2570289 (1.3)	2471873 (96.2)	2471718 (100)	37.68 (0.02)	0.91
ER0282	2776070 (1.1)	2631521 (94.8)	2631424 (100)	54.82 (0)	0.92
ER0307	2243071 (1.3)	2129896 (95)	2121283 (99.6)	65.92 (0.02)	0.84

Table 3.1. (cont'd)

ER0337	3254037 (1.4)	3104057 (95.4)	3104033 (100)	56.86 (0.99)	0.84
ER0412	1605913 (1.1)	1447033 (90.1)	1436481 (99.3)	44.6 (0.06)	0.77
ER0464	2496854 (1.6)	2381933 (95.4)	2381873 (100)	65.64 (0.02)	0.89
ER0491	3162424 (0.8)	3078034 (97.3)	3077926 (100)	61.01 (0.07)	0.84
ER0492	2231782 (1.2)	2113121 (94.7)	2112962 (100)	70.24 (0.06)	0.8
ER0515	2639718 (1.6)	2426865 (91.9)	2346267 (96.7)	74.26 (3.3)	0.84
ER0540	2379649 (1.1)	2249800 (94.5)	2249726 (100)	63.07 (2.22)	0.87
ER0550	3356714 (1.3)	3254937 (97)	3252732 (99.9)	58.53 (0.04)	0.9
ER0555	3562029 (1.2)	3458556 (97.1)	3458347 (100)	77.72 (0.01)	0.92
ER0560	3543311 (1.7)	3377653 (95.3)	3377371 (100)	56.23 (0)	0.94
ER0572	214000 (1.8)	203158 (94.9)	202618 (99.7)	58.79 (0)	0.52
ER0611	2430348 (1.7)	2301399 (94.7)	2300918 (100)	71.48 (0.01)	0.88
ER0624	2868691 (1.8)	2755771 (96.1)	2755499 (100)	71.56 (0.02)	0.88
ER0625	3723758 (1.2)	2502102 (67.2)	2501405 (100)	55.34 (0)	0.88
ER0647	2888229 (1.4)	2777506 (96.2)	2775514 (99.9)	69.25 (0.21)	0.85
ER0658	2570167 (1.9)	2454406 (95.5)	2454118 (100)	61.59 (0.33)	0.82
ER0663	2611261 (1.4)	2446519 (93.7)	2446423 (100)	74.71 (0)	0.88
ER0666	2958009 (1.3)	2819037 (95.3)	2818652 (100)	68.82 (0)	0.88
ER0670	2641744 (1.3)	2437352 (92.3)	2436914 (100)	69.92 (0.03)	0.93
ER0697	2673696 (1.5)	2397073 (89.7)	2396593 (100)	68.29 (0.04)	0.8
ER0710	2323990 (1.3)	2173969 (93.5)	2173724 (100)	70.73 (0.05)	0.86
ER0712	2706939 (1.3)	2597534 (96)	2597311 (100)	65.47 (0.01)	0.9
ER0740	3397680 (1.2)	3246871 (95.6)	3246679 (100)	61.59 (0.04)	0.88
ER0742	2660072 (1.4)	1948236 (73.2)	1947274 (100)	75.14 (0.3)	0.83
ER0762	2882882 (1.7)	2803803 (97.3)	2803576 (100)	76.18 (0.06)	0.9

Table 3.1. (cont'd)

ER0779	2289969 (1.3)	2097476 (91.6)	2095887 (99.9)	71.62 (0.04)	0.84
ER0795	3528372 (1.4)	3371056 (95.5)	3370768 (100)	71.88 (0.34)	0.91
ER0834	2931611 (1.1)	2473584 (84.4)	2469797 (99.8)	31.01 (0.01)	0.81
ER0835	212000 (1.8)	200843 (94.7)	200787 (100)	69.48 (0.01)	0.47
ER0870	2579674 (1.5)	2506714 (97.2)	2506116 (100)	72.46 (0)	0.91
ER0885	3281745 (1.6)	3141879 (95.7)	3141817 (100)	75.92 (0.04)	0.88
ER0948	3189071 (1.3)	3068392 (96.2)	3068309 (100)	70.66 (0.01)	0.91
ER0960	3145035 (1.6)	3026876 (96.2)	3026643 (100)	79.74 (0)	0.91
ER0973	2908116 (1.6)	2794297 (96.1)	2793590 (100)	39.5 (0.03)	0.85
ER0977	3169575 (1.6)	3065737 (96.7)	3064764 (100)	51.15 (0.01)	0.93
ER1007	2212195 (1.5)	2130732 (96.3)	2130353 (100)	65.66 (0.02)	0.73
ER1009	2144952 (1.6)	2079254 (96.9)	2078492 (100)	77.58 (0)	0.95
ER1020	2600216 (1.1)	2519880 (96.9)	2519823 (100)	57.78 (0.03)	0.83
ER0322	5205573 (1.1)	5097317 (97.9)	5097210 (100)	70.07 (0.03)	0.92
ER1008	5195123 (1.3)	5016461 (96.6)	5013802 (99.9)	53.02 (0.12)	0.88
ER1011	2618786 (1.7)	2450623 (93.6)	2450453 (100)	73.16 (0.01)	0.88

Table 3.2. Characteristics of the 79 patients with enteric infections and 63 recovered included in this study

Characteristic	No. of cases‡	Percentage (%) of cases	No. of follow up ‡	Percentage (%) of follow up	*p*-value
Demographic data					
Sex					
Male	38	48.1	28	44.4	-
Female	41	51.9	35	55.6	0.6642
Age group (years)					
0-9	21	26.6	17	27	0.9109
10-18	11	13.9	8	12.7	0.7872
19-64	33	41.8	26	41.3	0.8585
65+	14	17.7	12	19	-
Race					
Caucasian	60	81.1	50	87.7	0.5609
African American	10	13.5	5	8.8	1.0
Other	4	5.4	2	3.5	-
Residence Type					
Rural	40	51.3	33	53.2	0.8191
Urban	38	48.7	29	46.8	-
Residence (counties in Michigan)					
Calhoun	1	1.3	1	1.6	1.0
Clinton	4	5.1	4	6.5	0.6806
Eaton	5	6.4	5	8.1	0.4736
Ingham	16	20.5	11	17.7	0.7811
Ionia	2	2.6	1	1.6	1.0
Kent	5	6.4	4	6.5	0.7041
Lenawee	1	1.3	1	1.6	1.0
Livingston	3	3.9	3	4.8	0.6526
Macomb	3	3.9	3	4.8	0.6526
Oakland	8	10.2	7	11.3	0.5267
Ottawa	3	3.9	3	4.8	0.6526
Washtenaw	11	14.1	10	16.1	0.5504
Wayne	16	20.5	9	14.5	-
Infection					
Campylobacter	29	36.7	25	39.7	0.7192
Salmonella	35	44.3	28	44.4	1.0
Shigella	10	12.7	7	11.1	1.0
STEC	5	6.3	3	4.8	-
Epidemiological data					
Travel					
Domestic travel					
Yes	16	21.3	16	30.8	0.2284
No	59	78.7	36	69.2	-
International travel					
Yes	9	11.8	1	1.9	0.0476
No	67	88.2	51	98.1	-

Table 3.2. (cont'd)

Food consumption					
Turkey					
Yes	10	40	17	29.9	0.3667
No	15	60	40	70.1	-
Chicken					
Yes	55	84.6	54	94.7	0.0840
No	10	15.4	3	5.3	-
Beef					
Yes	39	88.6	48	84.2	0.5233
No	5	11.4	9	15.7	-
Pork*					
Yes	33	82.5	32	56.1	0.0065
No	7	17.5	25	43.9	-
Deli meat					
Yes	25	51	29	50.9	0.9882
No	24	49	28	49.1	-
Raw fruits					
Yes	31	83.8	51	89.5	0.4193
No	6	16.2	6	10.5	-
Raw leafy greens					
Yes	36	67.9	41	71.9	0.4191
No	17	32.1	16	28.1	-
Raw vegetables					
Yes	21	61.8	40	70.2	0.4089
No	13	38.2	17	29.8	-
Raw eggs					
Yes	1	2.5	2	3.5	1.0
No	39	97.5	55	96.5	-
Water at home					
Any well	13	18.8	5	11.4	1.0
Any municipal	48	69.6	26	74.3	1.0
Only bottled	8	11.6	4	14.3	-

The percentages based on the number for which information was available. Counts are mutually exclusive for each category. ‡ Total number varies due to the difference in missing data. * indicates significance difference ($p < 0.05$) between variables using *p*-value calculated by Chi-square test and Fisher's exact test for variables ≤ 5 in at least one cell. Mantel-Hanzel Chi-square was used to assess for trends.

Table 3.3. Characteristics of clusters defined through hierarchical clustering

Characteristic	Cluster 1‡ No. (%)	Cluster 2‡ No. (%)	Cluster 3‡ No. (%)	Cluster 4‡ No. (%)	*p*-value
Demographic data					
Case status					
Case*	25 (93)	33 (100)	7 (31.2)	14 (23.3)	0.0001
Follow Up	2 (7)	0 (0)	15 (68.2)	46 (76.7)	-
Sex					
Male	11 (40.7)	18 (54.5)	13 (59.1)	24 (40)	-
Female	16 (59.3)	15 (45.5)	9 (40.9)	36 (60)	0.3031
Age group (years)					
0-9	8 (29.6)	8 (24.2)	7 (31.8)	15 (25)	0.9212
10-18	3 (11.1)	6 (18.2)	3 (13.6)	7 (11.7)	0.9699
19-64	12 (44.4)	13 (39.4)	8 (36.4)	26 (43.3)	0.9585
65+	4 (14.8)	6 (18.2)	4 (18.2)	12 (20)	-
Race					
Caucasian	21 (87.5)	23 (76.7)	16 (72.7)	50 (90.9)	0.0085
African American	2 (8.3)	6 (20)	2 (9.1)	5 (9.1)	0.0713
Other	1 (4.2)	1 (3.3)	4 (18.2)	0 (0)	-
Residence Type					
Rural*	18 (66.7)	11 (33.3)	9 (40.9)	35 (60.3)	0.0219
Urban	9 (33.3)	22 (66.7)	13 (59.1)	23 (39.7)	-
Infection					
Campylobacter	8 (29.6)	11 (33.3)	9 (40.9)	26 (43.3)	0.7004
Salmonella	17 (63)	13 (39.4)	10 (45.5)	23 (38.3)	0.2486
Shigella	2 (7.4)	8 (24.2)	1 (4.5)	6 (10)	0.1458
STEC	0 (0)	1 (3)	2 (9.1)	5 (8.3)	-
Epidemiological data					
Travel					
Domestic travel					
Yes	8 (30.8)	8 (25.8)	3 (15.8)	13 (25.5)	0.7305
No	18 (69.2)	23 (74.2)	16 (84.2)	38 (74.5)	-
International travel					
Yes	4 (14.8)	2 (6.5)	2 (10.5)	2 (3.9)	0.0666
No	23 (85.2)	29 (93.5)	17 (89.5)	49 (96.1)	-
Food consumption					
Turkey					
Yes	5 (55.6)	3 (25)	4 (30.8)	15 (31.3)	0.5133
No	4 (44.4)	9 (75)	9 (69.2)	33 (68.8)	-
Chicken					
Yes	24 (88.9)	19 (82.6)	15 (93.8)	51 (91.1)	0.7514
No	3 (11.1)	4 (17.4)	1 (6.3)	5 (8.9)	-
Beef					
Yes	18 (100)	13 (81.3)	14 (87.5)	42 (82.4)	0.2522
No	0 (0)	3 (18.8)	2 (12.5)	9 (17.6)	-
Pork					
Yes	18 (94.7)	10 (66.7)	8 (53.3)	29 (60.4)	0.0175
No	1 (5.3)	5 (33.3)	7 (46.7)	19 (39.6)	-

Table 3.3. (cont'd)

Deli meat					
Yss	9 (42.9)	9 (47.4)	10 (62.5)	26 (52)	0.6767
No	12 (57.1)	10 (52.6)	6 (37.5)	24 (48)	-
Raw fruits					
Yes	10 (83.3)	12 (70.6)	14 (100)	46 (90.2)	0.0728
No	2 (16.7)	5 (29.4)	0 (0)	5 (9.8)	-
Raw leafy greens					
Yes	15 (65.2)	15 (71.4)	8 (50)	39 (78)	0.1837
No	8 (34.8)	6 (28.6)	8 (50)	11 (22)	-
Raw vegetables					
Yes	12 (70.6)	6 (42.9)	6 (46.2)	37 (78.7)	0.0707
No	5 (29.4)	8 (57.1)	7 (53.8)	10 (21.3)	-
Raw eggs					
Yes	0 (0)	1 (5.6)	0 (0)	2 (4.2)	1.0
No	18 (100)	17 (94.4)	13 (100)	46 (95.8)	-
Water at home					
Any well	2 (8.7)	4 (14.3)	3 (17.6)	9 (25)	0.7743
Any municipal	18 (78.3)	21 (75)	12 (70.6)	23 (63.9)	1.0
Only bottled	3 (13)	3 (10.7)	2 (11.8)	4 (11.1)	-

The percentages based on the number for which information was available. Counts are mutually exclusive for each category. ‡ Total number varies due to the difference in missing data. * indicates significance difference ($p < 0.05$) between variables using *p*-value calculated by Chi-square test and Fisher's exact test for variables ≤ 5 in at least one cell. Mantel-Hanzel Chi-square was used to assess for trends.

Table 3.4. Univariate analysis to identify disease associations for Cluster 1 in 79 patients with enteric infections included in the study

Characteristic	Totals*	No (%) Cluster 1	OR (95% CI)†	*p*-value‡
Sex				
Male	38	9 (16.7)	1.0	-
Female	41	16 (21.1)	0.5 (0.2 - 1.3)	0.1430
Age group (years)				
0-9	21	8 (38.1)	0.8 (0.23 - 3)	0.7753
10-18	11	2 (8)	2.2 (0.4 - 24.5)	0.4606
19-64	33	11 (44)	1.0	-
65+	14	4 (16)	1.2 (0.3 - 6.7)	1
Race				
Caucasian	60	21 (35)	0.6 (0.01 - 8.3)	0.6834
African American	10	2 (20)	1.3 (0.02 - 35)	0.8368
Other	4	1 (25)	1.0	-
Residence Type				
Urban	38	8 (21.1)	1.0	-
Rural	40	17 (42.5)	2.8 (1 - 7.5)	0.0425
Infection				
Camapylobacter	29	8 (27.6)	0 (0 - 3.6)	0.3086
Salmonella	35	16 (45.7)	0 (0 - 1.5)	0.0712
Shigella	10	1 (10)	0 (0 - 78)	1
STEC	5	0 (0)	1.0	-
Hospitalized				
Yes	29	6 (20.7)	0.4 (0.1 - 1.2)	0.0981
No	49	19 (38.8)	1.0	-
Abdominal pain				
Yes	65	21 (32.3)	1 (0.2 - 4.8)	1
No	12	4 (33.3)	1.0	-
Body ache				
Yes	22	14 (63.6)	7 (2.4 - 20.8)	0.0002
No	55	11 (20)	1.0	-
Diarrhea				
Yes	73	24 (32.9)	1.5 (0.1 - 80.3)	1
No	4	1 (25)	1.0	-
Bloody diarrhea				
Yes	29	7 (24.1)	0.5 (0.2 - 1.5)	0.2250
No	48	18 (37.5)	1.0	-
Chills				
Yes	25	10 (40)	1.6 (0.6 - 4.5)	0.3277
No	52	15 (28.8)	1.0	-
Fatigue				
Yes	41	14 (34.1)	1.2 (0.5 - 3.1)	0.7371
No	36	11 (30.6)	1.0	-

Table 3.4. (cont'd)

Headache				
Yes	18	6 (33.3)	1.1 (0.3 - 3.2)	0.9286
No	59	19 (32.2)	1.0	-
Nausea				
Yes	38	13 (34.2)	1.2 (0.5 - 3)	0.7471
No	39	12 (30.8)	1.0	-
Vomiting				
Yes	27	6 (22.2)	0.5 (0.2 - 1.4)	0.1583
No	50	19 (38)	1.0	-
Fever				
Yes	47	15 (31.9)	0.8 (0.3 - 2.2)	0.6187
No	21	8 (38.1)	1.0	-

* Depending on the variable examined, the number does not add up to the total (n=79) because of missing data. † 95% confidence interval (CI) for odds ratio (OR). ‡ *p*-value calculated by Chi-square test and Fisher's exact test was used for variables $\leq$5 in at least one cell. Mantel-Hanzel Chi-square was used to assess for trends.

Table 3.5. Univariate analysis to identify disease associations for Cluster 2 in 79 patients with enteric infections included in the study

Characteristic	Total*	No (%) Cluster 2	OR (95% CI)†	*p*-value‡
Sex				
Male	38	18 (47.4)	1.0	-
Female	41	15 (36.6)	1.6 (0.6 - 3.8)	0.3316
Age group (years)				
0-9	21	8 (38)	1.1 (0.3 - 3.3)	0.9240
10-18	11	6 (54.1)	0.5 (0.14 - 2.1)	0.3796
19-64	33	13 (39.4)	1.0	-
65+	14	6 (42.9)	0.9 (0.2 - 3.1)	0.8249
Race				
Caucasian	60	23 (38.3)	0.5 (0.01 - 7.2)	1
African American	10	6 (60)	0.2 (0.003 - 4)	0.5594
Other	4	1 (25)	1.0	-
Residence Type				
Urban	38	22 (57.9)	3.6 (1.4 - 9.3)	0.0112
Rural	40	11 (27.5)	1.0	-
Infection				
Camapylobacter	29	11 (37.9)	0.4 (0.01 - 5)	0.6347
Salmonella	35	13 (37.1)	0.4 (0.01-5)	0.6404
Shigella	10	8 (80)	0.1 (0.001-1.3)	0.0889
STEC	5	1 (25)	1.0	-
Hospitalized				
Yes	29	16 (55.2)	2.3 (0.9 - 5.9)	0.0769
No	49	17 (34.7)	1.0	-
Abdominal pain				
Yes	65	27 (41.5)	1.4 (0.3 - 7.1)	0.7525
No	12	4 (33.3)	1.0	-
Body ache				
Yes	22	6 (27.3)	0.4 (0.2 - 1.3)	0.1417
No	55	25 (45.5)	1.0	-
Diarrhea				
Yes	73	31 (42.5)	Un (0.5 - Un)	0.1438
No	4	0 (0)	1.0	-
Bloody diarrhea				
Yes	29	15 (51.7)	2.1 (0.8 - 5.5)	0.1108
No	48	16 (33.3)	1.0	-
Chills				
Yes	25	8 (32)	0.6 (0.2 - 1.6)	0.3055
No	52	23 (44.2)	1.0	-
Fatigue				
Yes	41	15 (36.6)	0.7 (0.3 - 1.8)	0.4829
No	36	16 (44.4)	1.0	-

Table 3.5. (cont'd)

Headache				
Yes	18	10 (55.6)	2.3 (0.8 - 6.6)	0.1306
No	59	21 (35.6)	1.0	-
Nausea				
Yes	38	15 (39.5)	0.9 (0.4 - 2.3)	0.8896
No	39	16 (41)	1.0	-
Vomiting				
Yes	27	18 (66.7)	5.7 (2.1 - 15.8)	0.0005
No	50	13 (26)	1.0	-
Fever				
Yes	47	21 (44.7)	2 (0.7 - 6.1)	0.2097
No	21	6 (28.6)	1.0	-

* Depending on the variable examined, the number does not add up to the total (n=79) because of missing data. † 95% confidence interval (CI) for odds ratio (OR). ‡ *p*-value calculated by Chi-square test and Fisher's exact test was used for variables $\leq$5 in at least one cell. Mantel-Hanzel Chi-square was used to assess for trends.

Table 3.6. Differentially abundant taxa determined by ANCOM for each case cluster

Organism (Genus)	Taxonomy (Order; Family)	Cluster 1	Cluster 2
Viruses			
P22virus	Caudovirales; Podoviridae		Present
P2virus	Caudovirales; Myoviridae		Present
Nona33virus	Caudovirales; Podoviridae		Present
Mastadenovirus	Viruses; Adenoviridae		Present
Lambdavirus	Caudovirales; Siphoviridae		Present
Orthopoxvirus	Viruses; Poxviridae		Present
Kp15virus	Caudovirales; Myoviridae		Present
P1virus	Caudovirales; Myoviridae		Present
T7virus	Caudovirales; Podoviridae		Present
C2virus	Caudovirales; Siphoviridae		Present
Phi29virus	Caudovirales; Podoviridae		Present
Sk1virus	Caudovirales; Siphoviridae		Present
Felixo1virus	Caudovirales; Myoviridae		Present
Epsilon15virus	Caudovirales; Podoviridae		Present
Jerseyvirus	Caudovirales; Siphoviridae		Present
V5virus	Caudovirales; Myoviridae		Present
T5virus	Caudovirales; Siphoviridae		Present
Sfi11virus	Caudovirales; Siphoviridae		Present
Pis4avirus	Caudovirales; Siphoviridae		Present
Muvirus	Caudovirales; Myoviridae		Present
Sfi21dt1virus	Caudovirales; Siphoviridae		Present
K1gvirus	Caudovirales; Siphoviridae		Present
Kayvirus	Caudovirales; Herelleviridae		Present
Cytomegalovirus	Herpesvirales; Herpesviridae		Present
Tl2011virus	Caudovirales; Podoviridae		Present
Hk578virus	Caudovirales; Siphoviridae		Present
Rb69virus	Caudovirales; Myoviridae		Present
Seuratvirus	Caudovirales; Siphoviridae		Present
Np1viru	Caudovirales; Siphoviridae		Present
Bacteria			
Salmonella	Enterobacterales; Enterobacteriaceae	Present	Present
Escherichia	Enterobacterales; Enterobacteriaceae		Present
Clostridium	Clostridiales; Clostridiaceae		Present
Roseburia	Clostridiales; Lachnospiraceae		Present
Shigella	Enterobacterales; Enterobacteriaceae		Present
Blautia	Clostridiales; Lachnospiraceae		Present
Clostridioides	Clostridiales; Peptostreptococcaceae		Present
Klebsiella	Enterobacterales; Enterobacteriaceae		Present
Ruminococcus	Clostridiales; Ruminococcaceae		Present
Enterobacter	Enterobacterales; Enterobacteriaceae	Present	Present
Butyricicoccus	Clostridiales; Clostridiaceae		Present
Citrobacter	Enterobacterales; Enterobacteriaceae	Present	Present
Chlamydia	Chlamydiales; Chlamydiaceae		Present

Table 3.6. (cont'd)

Eubacterium	Clostridiales; Eubacteriaceae		Present
Lactococcus	Lactobacillales; Streptococcaceae		Present
Streptococcus	Lactobacillales; Streptococcaceae		Present
Flavonifractor	Clostridiales; Ruminococcaceae		Present
Haemophilus	Enterobacterales; Enterobacteriaceae	Present	Present
Vibrio	Vibrionales; Vibrionaceae		Present
Subdoligranulum	Clostridiales; Ruminococcaceae		Present
Anaerotruncus	Clostridiales; Ruminococcaceae		Present
Pantoea	Enterobacterales; Erwiniaceae	Present	Present
Coprococcus	Clostridiales; Lachnospiraceae		Present
Agathobaculum	Clostridiales; Ruminococcaceae		Present
Fusicatenibacter	Clostridiales; Lachnospiraceae		Present
Acinetobacter	Pseudomonadales; Moraxellaceae		Present
Prevotellamassilia	Bacteroidales; Prevotellaceae		Present
Pseudomonas	Pseudomonadales; Pseudomonadaceae		Present
Pseudoflavonifractor	Clostridiales; Ruminococcaceae		Present
Staphylococcus	Bacillales; Staphylococcaceae		Present
Oscillibacter	Clostridiales; Oscillospiraceae		Present
Serratia	Enterobacterales; Yersiniaceae		Present
Bacillus	Bacillales; Bacillaceae		Present
Enterococcus	Lactobacillales; Enterococcaceae		Present
Lactobacillus	Lactobacillales; Lactobacillaceae		Present
Alloprevotella	Bacteroidales; Prevotellaceae		Present
Anaerotignum	Clostridiales; Lachnospiraceae		Present
Intestinibacillus	Clostridiales; Eubacteriaceae		Present
Intestinimonas	Clostridiales; unclassified Clostridiales		Present
Ruthenibacterium	Clostridiales; Ruminococcaceae		Present
Atlantibacter	Enterobacterales; Enterobacteriaceae		Present
Butyrivibrio	Clostridiales; Lachnospiraceae		Present
Raoultella	Clostridiales; Lachnospiraceae	Present	Present
Gemmiger	Clostridiales; Ruminococcaceae		Present
Anaeromassilibacillus	Clostridiales; Ruminococcaceae		Present
Duodenibacillus	Burkholderiales; Sutterellaceae		Present
Kluyvera	Enterobacterales; Enterobacteriaceae	Present	Present
Angelakisella	Clostridiales; Ruminococcaceae		Present
Lawsonibacter	Clostridiales; unclassified Clostridiales		Present
Drancourtella	Clostridiales; Ruminococcaceae		Present
Peptostreptococcus	Clostridiales; Peptostreptococcaceae		Present
Proteus	Enterobacterales; Morganellaceae		Present
Synergistes	Synergistales; Synergistaceae		Present
Acetobacter	Rhodospirillales; Acetobacteraceae		Present
Hafnia	Enterobacterales; Hafniaceae		Present
Cloacibacillus	Synergistales; Synergistaceae		Present
Christensenella	Clostridiales; Christensenellaceae		Present

Table 3.6. (cont'd)

Providencia	Enterobacterales; Morganellaceae	Present
Neglecta	Clostridiales; Ruminococcaceae	Present
Morganella	Enterobacterales; Morganellaceae	Present
Colibacter	Veillonellales; Veillonellaceae	Present
Tissierella	Tissierellales; Tissierellaceae	Present
Culturomica	Bacteroidales; Odoribacteraceae	Present

Table 3.7. Univariate and multivariate analysis of microbial factors for Cluster 2 status in 79 patients with enteric infections and 63 recovered included in the study

Characteristic	Total*	No (%) Cluster 2	OR (95% CI) †	*p*-value‡
Viruses above study average				
P22virus				
Yes	11	6 (54.5)	4.6 (1.1 - 20.4)	0.0196
No	131	27 (20.6)	1.0	-
P2virus				
Yes	13	7 (53.8)	4.6 (1.4 - 14.9)	0.0061
No	129	26 (20.2)	1.0	-
Nona33virus				
Yes	18	13 (72.2)	13.2 (3.9 - 52.5)	< 0.0001
No	124	20 (16.1)	1.0	-
Mastadenovirus				
Yes	1	1 (100)	Inf (0.1 - Inf)	0.2324
No	141	32 (22.7)	1.0	-
Lambdavirus				
Yes	15	10 (66.7)	8.8 (2.5 - 36.3)	0.0002
No	127	23 (18.1)	1.0	-
Orthopoxvirus				
Yes	30	20 (66.7)	15.2 (5.9 - 39.5)	< 0.0001
No	112	13 (11.6)	1.0	-
Kp15virus				
Yes	1	1 (100)	Inf (0.1 - Inf)	0.2324
No	141	32 (22.7)	1.0	-
P1virus				
Yes	11	4 (36.4)	2 (0.4 - 8.5)	0.281
No	131	29 (22.1)	1.0	-
T7virus				
Yes	10	1 (10)	0.3 (0 - 2.7)	0.4533
No	132	32 (24.2)	1.0	-
C2virus				
Yes	3	1 (33.3)	1.7 (0 - 33)	0.5506
No	139	32 (23)	1.0	-
Phi29virus				
Yes	13	2 (15.4)	0.6 (0.1 - 2.9)	0.7327
No	129	31 (24)	1.0	-
Sk1virus				
Yes	7	1 (14.3)	0.5 (0 - 4.7)	1
No	135	32 (23.7)	1.0	-
Felixo1virus				
Yes	2	1 (50)	3.3 (0 - 266.7)	0.412
No	140	32 (22.9)	1.0	-

Table 3.7. (cont'd)

Epsilon15virus				
Yes	9	4 (44.4)	2.8 (0.5 - 14.2)	0.2127
No	133	29 (21.8)	1.0	-
Jerseyvirus				
Yes	4	1 (25)	1.1 (0 - 14.3)	1
No	138	32 (23.2)	1.0	-
V5virus				
Yes	2	1 (50)	3.3 (0 - 266.7)	0.412
No	140	32 (22.9)	1.0	-
T5virus				
Yes	2	2 (100)	Inf (0.6 - Inf)	0.0527
No	140	31 (22.1)	1.0	-
Sfi11virus				
Yes	12	0 (0)	0 (0 - 1.1)	0.0685
No	130	33 (25.4)	1.0	-
Pis4avirus				
Yes	4	3 (75)	10.6 (0.8 - 570)	0.0392
No	138	30 (21.7)	1.0	-
Muvirus				
Yes	5	3 (60)	5.3 (0.6 - 65.8)	0.0822
No	137	30 (21.9)	1.0	-
Sfi21dt1virus				
Yes	12	0 (0)	0 (0 - 1.1)	0.0685
No	130	33 (25.4)	1.0	-
K1gvirus				
Yes	4	1 (25)	1.1 (0 - 14.3)	1
No	138	32 (23.2)	1.0	-
Kayvirus				
Yes	19	8 (42.1)	2.9 (1 - 7.8)	0.0364
No	123	25 (20.3)	1.0	-
Cytomegalovirus				
Yes	11	8 (72.7)	11 (2.4 - 69.1)	0.0004
No	131	25 (19.1)	1.0	-
Tl2011virus				
Yes	7	3 (42.9)	2.6 (0.4 - 16.3)	0.3535
No	135	30 (22.2)	1.0	-
Hk578virus				
Yes	5	1 (20)	0.8 (0 - 8.7)	1
No	137	32 (23.4)	1.0	-
Rb69virus				
Yes	1	0 (0)	0 (0 - 128.6)	1
No	141	33 (23.4)	1.0	-

Table 3.7. (cont'd)

Seuratvirus				
Yes	10	6 (60)	5.7 (1.3 - 29.7)	0.0108
No	132	27 (20.5)	1.0	-
Np1virus				
Yes	14	8 (57.1)	5.5 (1.7 - 17.3)	0.0016
No	128	25 (19.5)	1.0	-
Bacteria above study average				
Salmonella				
Yes	17	8 (47.1)	3.6 (1.2 - 10.1)	0.0132
No	125	25 (20)	1.0	-
Escherichia				
Yes	33	21 (63.6)	14.1 (5.6 - 35.8)	< 0.0001
No	109	12 (11)	1.0	-
Clostridium				
Yes	40	2 (5)	0.1 (0 - 0.5)	0.0008
No	102	31 (30.4)	1.0	-
Roseburia				
Yes	28	0 (0)	0 (0 - 0.4)	0.0003
No	114	33 (28.9)	1.0	-
Shigella				
Yes	31	22 (71)	22.2 (8.2 - 60.1)	< 0.0001
No	111	11 (9.9)	1.0	-
Blautia				
Yes	31	0 (0)	0 (0 - 0.3)	0.0002
No	111	33 (29.7)	1.0	-
Clostridioides				
Yes	51	7 (13.7)	0.4 (0.2 - 1)	0.0445
No	91	26 (28.6)	1.0	-
Klebsiella				
Yes	19	10 (52.6)	4.8 (1.8 - 13.2)	0.0011
No	123	23 (18.7)	1.0	-
Ruminococcus				
Yes	46	0 (0)	0 (0 - 0.2)	< 0.0001
No	96	33 (34.4)	1.0	-
Enterobacter				
Yes	25	20 (80)	30.6 (9.3 - 123.5)	< 0.0001
No	117	13 (11.1)	1.0	-
Butyricicoccus				
Yes	34	0 (0)	0 (0 - 0.3)	< 0.0001
No	108	33 (30.6)	1.0	-
Citrobacter				
Yes	19	9 (47.4)	3.7 (1.4 - 10.1)	0.0075
No	123	24 (19.5)	1.0	-

Table 3.7. (cont'd)

Chlamydia				
Yes	36	11 (30.6)	1.7 (0.7 - 3.9)	0.229
No	106	22 (20.8)	1.0	-
Eubacterium				
Yes	33	0 (0)	0 (0 - 0.3)	0.0001
No	109	33 (30.3)	1.0	-
Lactococcus				
Yes	44	7 (15.9)	0.5 (0.2 - 1.3)	0.1658
No	98	26 (26.5)	1.0	-
Streptococcus				
Yes	20	10 (50)	4.3 (1.6 - 11.5)	0.0022
No	122	23 (18.9)	1.0	-
Flavonifractor				
Yes	34	1 (2.9)	0.1 (0 - 0.5)	0.0008
No	108	32 (29.6)	1.0	-
Haemophilus				
Yes	20	10 (50)	4.3 (1.6 - 11.5)	0.0022
No	122	23 (18.9)	1.0	-
Vibrio				
Yes	19	13 (68.4)	11.2 (3.8 - 32.8)	< 0.0001
No	123	20 (16.3)	1.0	-
Subdoligranulum				
Yes	40	0 (0)	0 (0 - 0.2)	< 0.0001
No	102	33 (32.4)	1.0	-
Anaerotruncus				
Yes	31	0 (0)	0 (0 - 0.3)	0.0002
No	111	33 (29.7)	1.0	-
Pantoea				
Yes	46	32 (69.6)	204 (29.7 - 8376.8)	< 0.0001
No	96	1 (1)	1.0	-
Coprococcus				
Yes	37	2 (5.4)	0.1 (0 - 0.6)	0.0027
No	105	31 (29.5)	1.0	-
Agathobaculum				
Yes	32	0 (0)	0 (0 - 0.3)	0.0001
No	110	33 (30)	1.0	-
Fusicatenibacter				
Yes	38	1 (2.6)	0.1 (0 - 0.4)	0.0002
No	104	32 (30.8)	1.0	-
Acinetobacter				
Yes	16	14 (87.5)	37.9 (7.8 - 369.5)	< 0.0001
No	126	19 (15.1)	1.0	-

Table 3.7. (cont'd)

Prevotellamassilia				
Yes	9	2 (22.2)	0.9 (0.1 - 5.3)	1
No	133	31 (23.3)	1.0	-
Pseudomonas				
Yes	8	5 (62.5)	6.2 (1.1 - 42.4)	0.0169
No	134	28 (20.9)	1.0	-
Pseudoflavonifractor				
Yes	34	1 (2.9)	0.1 (0 - 0.5)	0.0008
No	108	32 (29.6)	1.0	-
Staphylococcus				
Yes	19	14 (73.7)	14.9 (4.4 - 59.3)	< 0.0001
No	123	19 (15.4)	1.0	-
Oscillibacter				
Yes	34	1 (2.9)	0.1 (0 - 0.5)	0.0008
No	108	32 (29.6)	1.0	-
Serratia				
Yes	19	16 (84.2)	31.9 (8 - 188.9)	< 0.0001
No	123	17 (13.8)	1.0	-
Bacillus				
Yes	33	13 (39.4)	2.9 (1.2 - 6.8)	0.0121
No	109	20 (18.3)	1.0	-
Enterococcus				
Yes	18	14 (77.8)	18.7 (5.2 - 86.5)	< 0.0001
No	124	19 (15.3)	1.0	-
Lactobacillus				
Yes	20	5 (25)	1.1 (0.3 - 3.6)	0.7824
No	122	28 (23)	1.0	-
Alloprevotella				
Yes	12	5 (41.7)	2.6 (0.6 - 10.3)	0.1498
No	130	28 (21.5)	1.0	-
Anaerotignum				
Yes	30	0 (0)	0 (0 - 0.3)	0.0002
No	112	33 (29.5)	1.0	-
Intestinibacillus				
Yes	33	1 (3)	0.1 (0 - 0.5)	0.0008
No	109	32 (29.4)	1.0	-
Intestinimonas				
Yes	32	0 (0)	0 (0 - 0.3)	0.0001
No	110	33 (30)	1.0	-
Ruthenibacterium				
Yes	27	0 (0)	0 (0 - 0.4)	0.0006
No	115	33 (28.7)	1.0	-

Table 3.7. (cont'd)

Atlantibacter				
Yes	3	1 (33.3)	1.7 (0 - 33)	0.5506
No	139	32 (23)	1.0	-
Butyrivibrio				
Yes	38	3 (7.9)	0.2 (0 - 0.8)	0.0075
No	104	30 (28.8)	1.0	-
Raoultella				
Yes	8	5 (62.5)	6.2 (1.1 - 42.4)	0.0169
No	134	28 (20.9)	1.0	-
Gemmiger				
Yes	34	0 (0)	0 (0 - 0.3)	< 0.0001
No	108	33 (30.6)	1.0	-
Anaeromassilibacillus				
Yes	19	0 (0)	0 (0 - 0.6)	0.007
No	123	33 (26.8)	1.0	-
Duodenibacillus				
Yes	7	2 (28.6)	1.3 (0.1 - 8.7)	0.6638
No	135	31 (23)	1.0	-
Kluyvera				
Yes	4	1 (25)	1.1 (0 - 14.3)	1
No	138	32 (23.2)	1.0	-
Angelakisella				
Yes	26	0 (0)	0 (0 - 0.4)	0.0006
No	116	33 (28.4)	1.0	-
Lawsonibacter				
Yes	34	0 (0)	0 (0 - 0.3)	< 0.0001
No	108	33 (30.6)	1.0	-
Drancourtella				
Yes	21	0 (0)	0 (0 - 0.6)	0.0039
No	121	33 (27.3)	1.0	-
Peptostreptococcus				
Yes	9	7 (77.8)	14 (2.5 - 146.1)	0.0005
No	133	26 (19.5)	1.0	-
Proteus				
Yes	2	1 (50)	3.3 (0 - 266.7)	0.412
No	140	32 (22.9)	1.0	-
Synergistes				
Yes	2	2 (100)	Inf (0.6 - Inf)	0.0527
No	140	31 (22.1)	1.0	-
Acetobacter				
Yes	11	1 (9.1)	0.3 (0 - 2.3)	0.4575
No	131	32 (24.4)	1.0	-

Table 3.7. (cont'd)

Hafnia				
Yes	4	3 (75)	10.6 (0.8 - 570)	0.0392
No	138	30 (21.7)	1.0	-
Cloacibacillus				
Yes	6	2 (33.3)	1.7 (0.1 - 12.4)	0.6232
No	136	31 (22.8)	1.0	-
Christensenella				
Yes	30	1 (3.3)	0.1 (0 - 0.6)	0.0028
No	112	32 (28.6)	1.0	-
Providencia				
Yes	6	3 (50)	3.5 (0.4 - 27.5)	0.1381
No	136	30 (22.1)	1.0	-
Neglecta				
Yes	19	0 (0)	0 (0 - 0.6)	0.007
No	123	33 (26.8)	1.0	-
Morganella				
Yes	3	1 (33.3)	1.7 (0 - 33)	0.5506
No	139	32 (23)	1.0	-
Colibacter				
Yes	2	1 (50)	3.3 (0 - 266.7)	0.412
No	140	32 (22.9)	1.0	-
Tissierella				
Yes	2	0 (0)	0 (0 - 17.7)	1
No	140	33 (23.6)	1.0	-
Culturomica				
Yes	2	0 (0)	0 (0 - 17.7)	1
No	140	33 (23.6)	1.0	-

	Multivariate Analysis		
Logistic Regression	**OR**	**95% CI €**	***p* value‡**
Model 1			
Pantoea			
Above study average: Yes	217.1	27.5 - 1717.2	< 0.0001
Model 2			
Pantoea			
Above study average: Yes	142.1	17.5 - 11157.2	< 0.0001
Serratia			
Above study average: Yes	10.7	1.6 - 70.5	0.01499
Model 3			
Serratia			
Above study average: Yes	10.7	1.6 - 70.5	0.01499
Pantoea			
Above study average: Yes	142.1	17.5 - 11157.2	< 0.0001

Table 3.7. (cont'd)

Model 4			
Pantoea			
Above study average: Yes	92.8	11.2 - 770.2	< 0.0001
Serratia			
Above study average: Yes	6.5	0.8 - 50.3	0.0758
Enterobacter			
Above study average: Yes	5.5	1.1 - 28.4	0.0425
Model 5			
Acinetobacter			
Above study average: Yes	4.9	0.6 - 43.6	0.1532
Enterobacter			
Above study average: Yes	5.8	1.1 - 29.3	0.0364
Pantoea			
Above study average: Yes	96.8	11.7 – 800.2	< 0.0001
Model 6			
Salmonella			
Above study average: Yes	5.2	1.2 - 22.4	0.1588
Enterobacter			
Above study average: Yes	9.1	1.8 - 45.9	0.0052
Pantoea			
Above study average: Yes	62.7	7.2 - 544.4	< 0.0001
Model 7			
Escherichia			
Above study average: Yes	3.3	0.8 - 13.2	0.1020
Enterobacter			
Above study average: Yes	7.4	1.5 - 35.9	0.0146
Pantoea			
Above study average: Yes	82.8	9.8 - 698.5	< 0.0001
Model 8			
Orthopoxvirus			
Above study average: Yes	3.1	0.8 - 13	0.1151
Enterobacter			
Above study average: Yes	7.8	1.6 - 37.4	0.0118
Pantoea			
Above study average: Yes	81.5	9.6 - 693.4	< 0.0001
Model 9			
Nona33virus			
Above study average: Yes	4.9	0.7 - 36.1	0.1192
Enterobacter			
Above study average: Yes	7	1.4 - 34	0.0181
Pantoea			
Above study average: Yes	103	12.5 - 851	< 0.0001

Table 3.7. (cont'd)

Model 10			
Shigella			
Above study average: Yes	5.2	1.2 - 22.4	0.0003
Enterobacter			
Above study average: Yes	9.1	1.8 - 45.9	0.0086
Pantoea			
Above study average: Yes	62.7	7.2 - 544.4	0.0004

	Model Performance		
	Accuracy	**Accuracy 95% CI**	**AUC**
Final Model (Model 10)	0.9722	(0.8547, 0.9993)	0.9955

* The number of isolates may not add up to the total (n=142) due to missing data.

† 95% confidence interval (CI) for odds ratio (OR)

‡ *p*-value was calculated by Chi-square test, and Fisher's exact test was used for variables ≤5 in at least one of the cells.

£ Logistic regression was performed via forward selection while controlling for variables that yielded strong ($p \leq 0.20$) associations with the outcome as Cluster 2 in the univariate analysis. Hosmer-Lemeshow Goodness-of-Fit test. All variables were tested for collinearity.

€ Wald 95% confidence intervals (CI)

Figure 3.1. Power analysis for chi-square and logistic regression modeling Power curves were created based on the Cohen power equations. The below curves show the relationship between the effect size (differences in means over pooled standard deviations) and the sample size needed to detect that effect size. The circle represents the study (n=142) within the 0.8 power curve (blue).

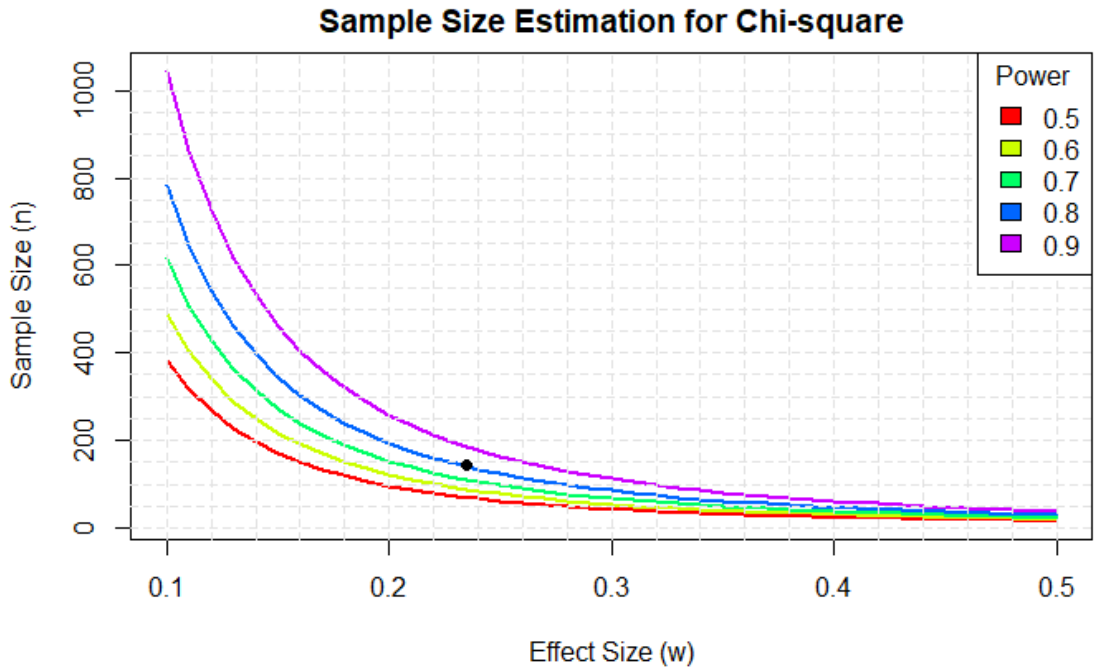


Figure 3.2. The percentage of reads annotated at four taxonomical levels The number of quality-controlled reads that were annotated compared to the total number of quality-controlled reads (n=142). The line in the box represents the median, and the interquartile range (25%-75%) is represented by the box. The whiskers are the confidence interval (5%-95%). Outliers represented as circles.

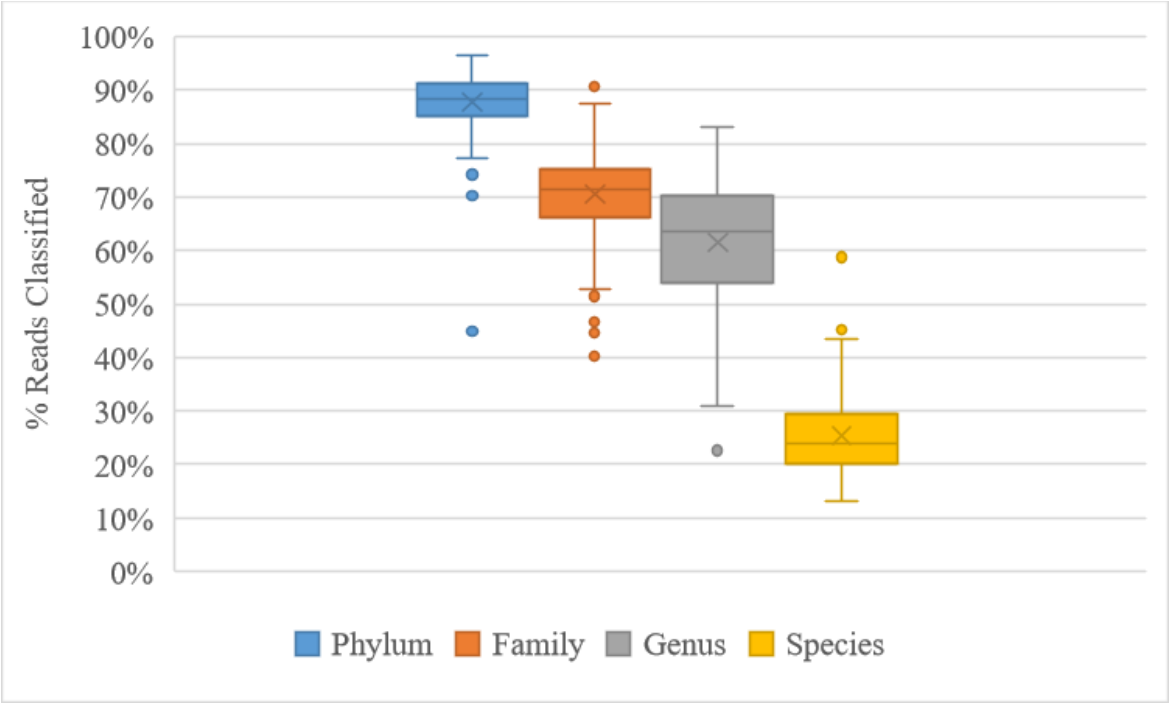


Figure 3.3. Rarefaction curves A) Random sampling assessed cumulative sequencing across all samples by study group and B) Rarefaction of total reads to assess the richness of genera. Curves represent plots of either case (red, n=79) or follow-up (purple, n=63) samples with confidence intervals (95%).

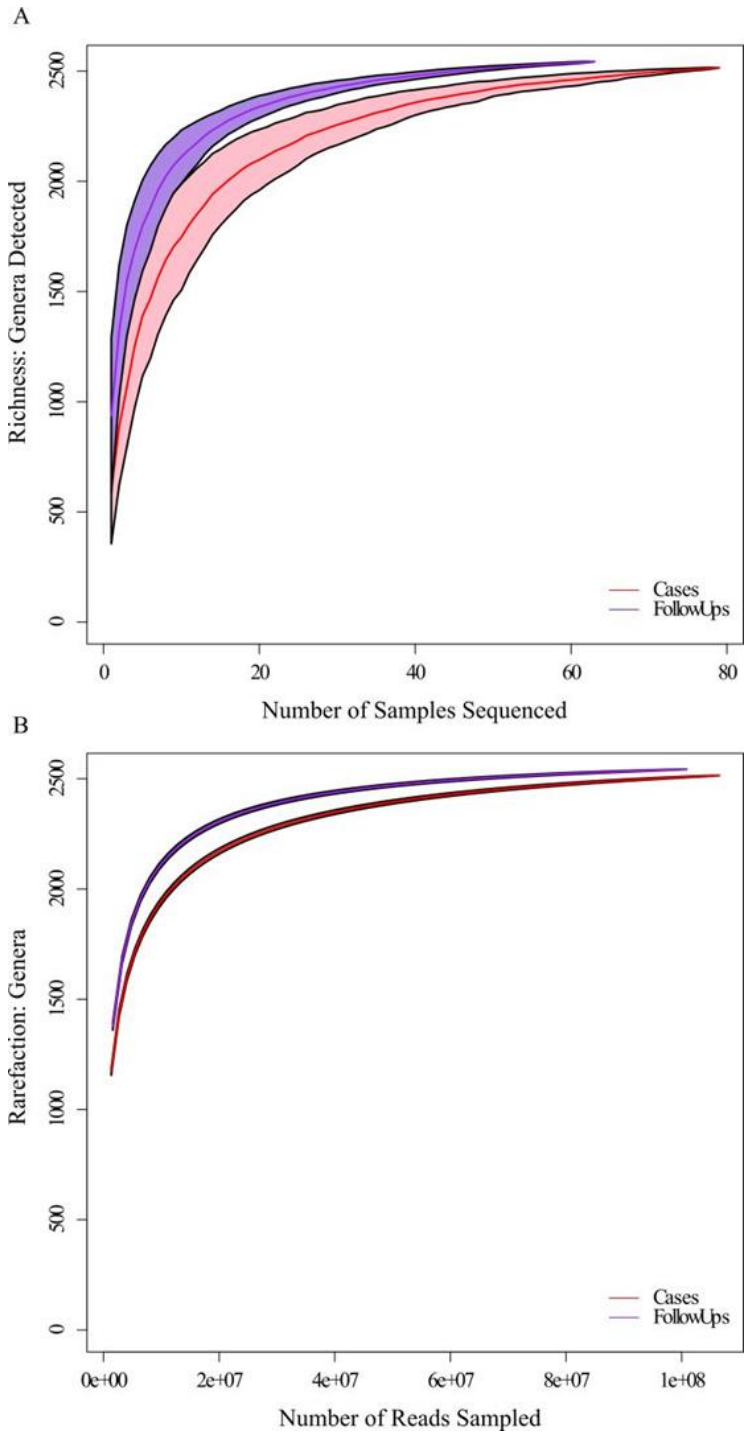


Figure 3.4 Diversity Metrics for the samples from 79 cases and 63 Followups. A) Shannon Index, B) Genera Richness, C) Evenness between groups. The lines in the boxes represent medians; the box is the interquartile range (25%-75%) and the whiskers are confidence intervals (95%). Outliers are circles. The asterisk (*) is significant finding ($p < 0.05$).

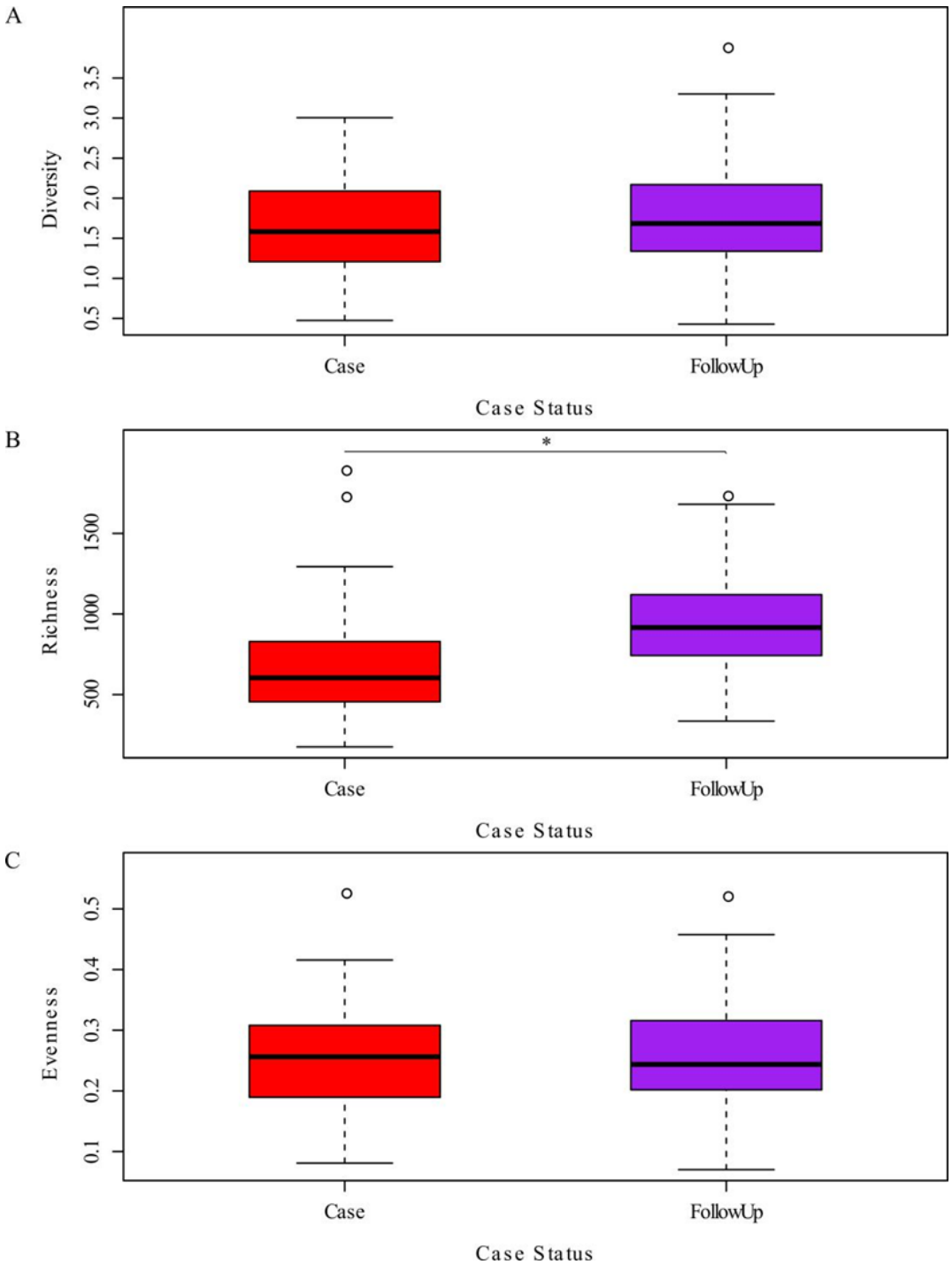


Figure 3.5. Microbiome profiles of patients during infection (Case) and post-recovery (FollowUp) A) The top 5 most abundant viruses across samples, and B) The top 10 most abundant bacteria. Both viruses and bacteria are presented at the Family taxonomical level. The line in the box represents the median. The interquartile range (25%-75%) in the box surrounding the median. The whiskers extend from 5%-95%. Outliers are circles.

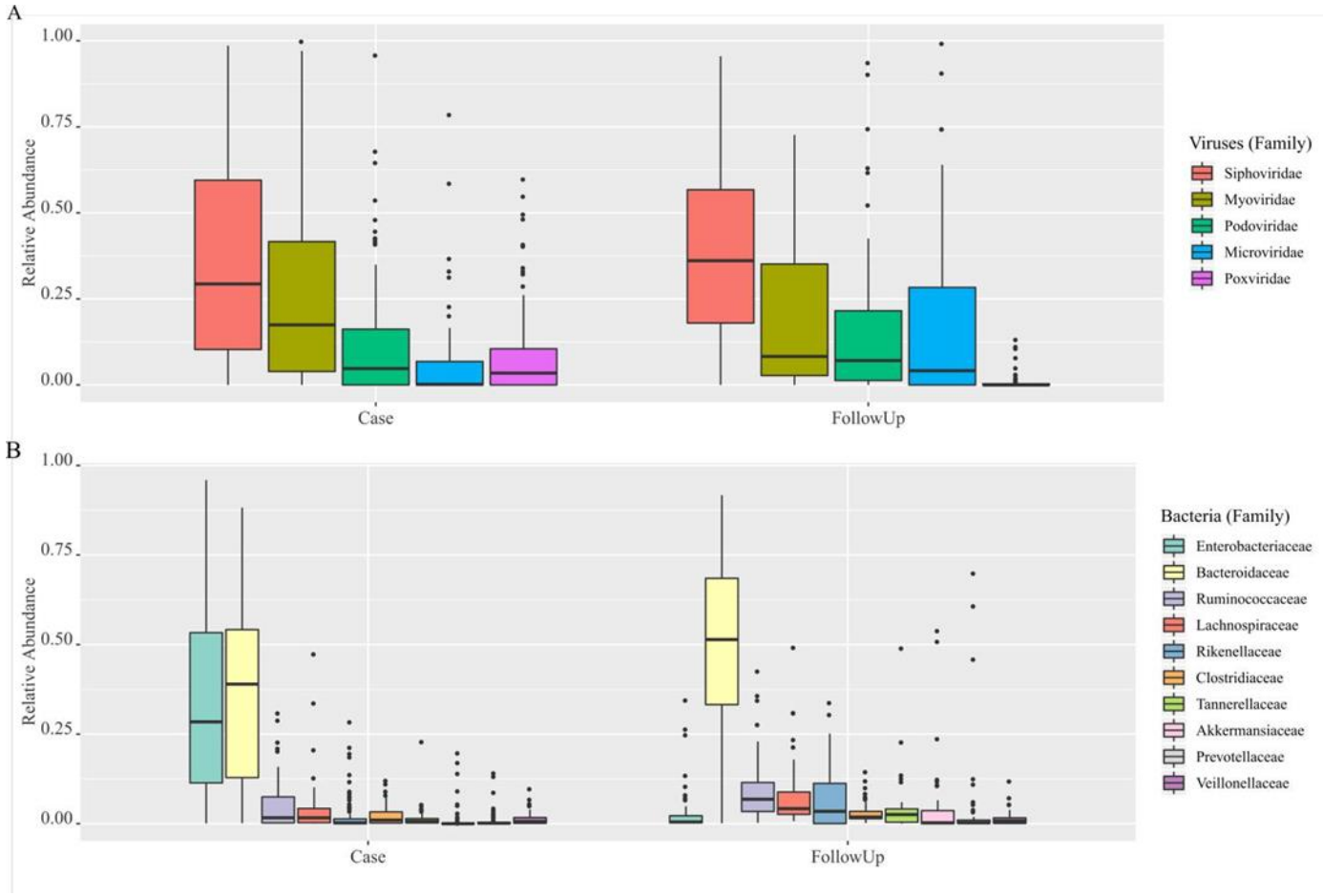


Figure 3.6. Microbiome clusters identified by hierarchical clustering A) In total, four distinct clusters were identified. Cases are represented with circles, and follow-ups samples are triangles. B) The beta-dispersion or heterogeneity of each cluster and the spatial relationship between points. The ellipses for both plots represent the 95% confidence intervals. Colors represent each cluster are the same for both panels.

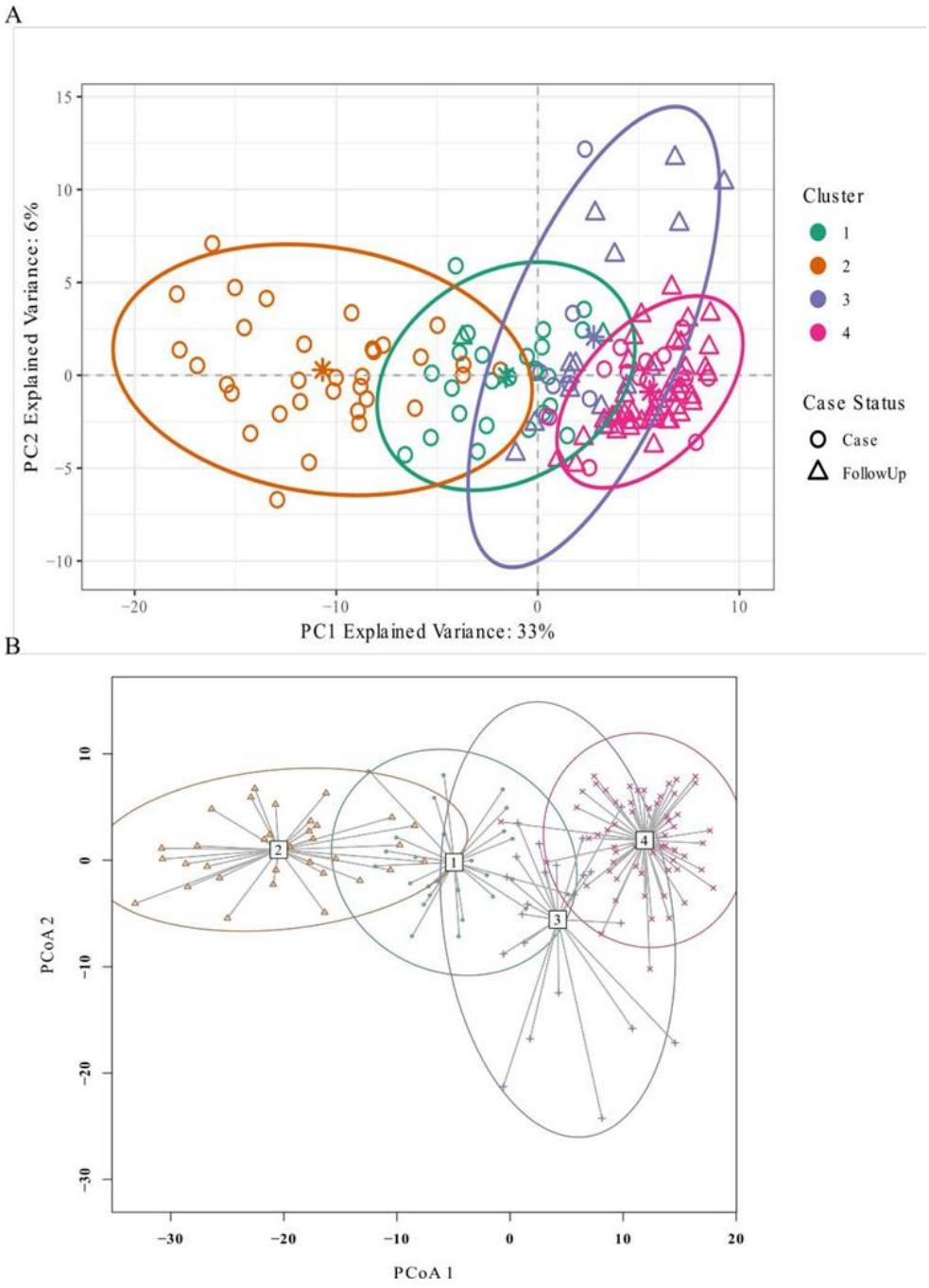


Figure 3.7. Diversity metrics for the microbiome profiles representing the four clusters A) Shannon Index, B) Genera Richness, C) Evenness between groups. The medians are represented the lines in the boxes, the box is the interquartile range (25%-75%), and the whiskers are the confidence intervals (95%). The asterisk (*) is significant finding ($p < 0.05$). Circles are outliers.

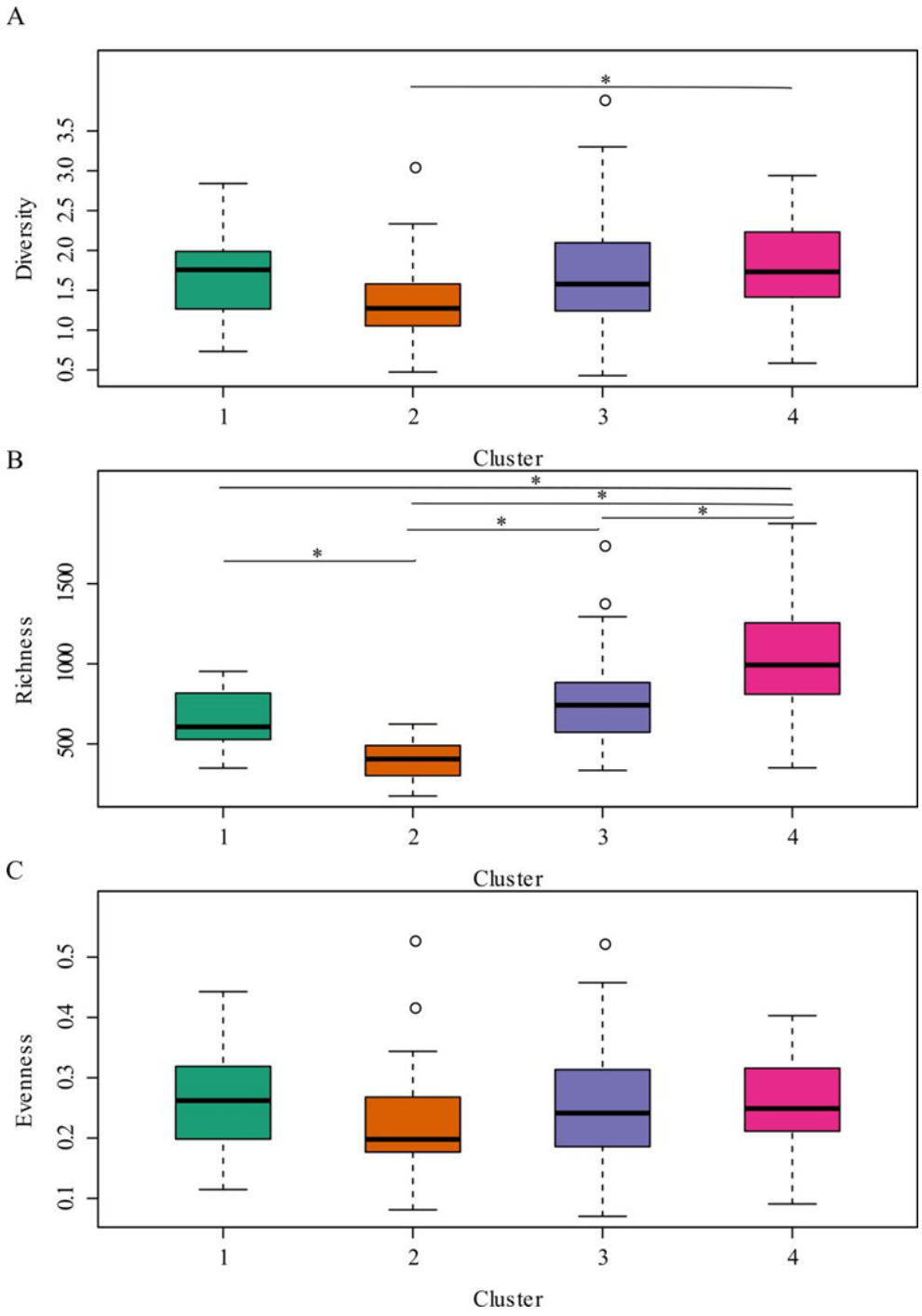


Figure 3.8. Community composition among samples representing the four clusters Clusters are colored to the previous PCA (cluster 1 = green, cluster 2 = orange, cluster 3 = purple, cluster 4 = pink). Rows are colored by cluster based on genera abundance. A dendrogram shows the clustering of the samples (top), genera (rows). The heatmap cell colors represent the number of standard deviations a genus is from the mean within a column. Purple is more abundant taxa genera, whereas orange coloring represents lower abundant genera within a sample.

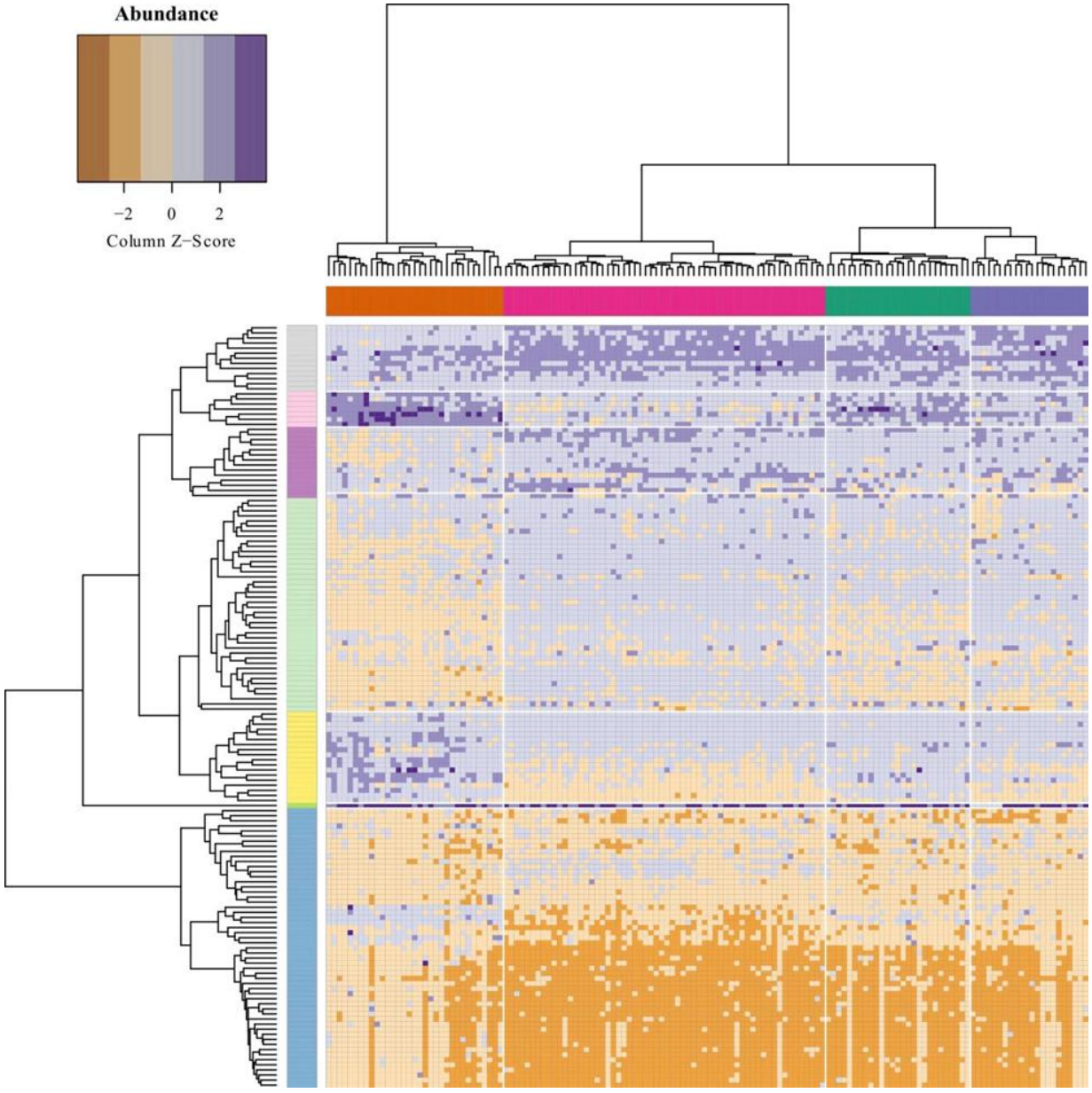


Figure 3.9. A network of differentially abundant microbes within Cluster 2 communities

The vertices represent taxa and named by genus. The size of the vertex represents the abundance found across samples and are colored by higher taxonomical classification. Significant correlations are represented (absolute value ≥ 0.3) by the edges; positive correlations are green, and negative correlations are pink.

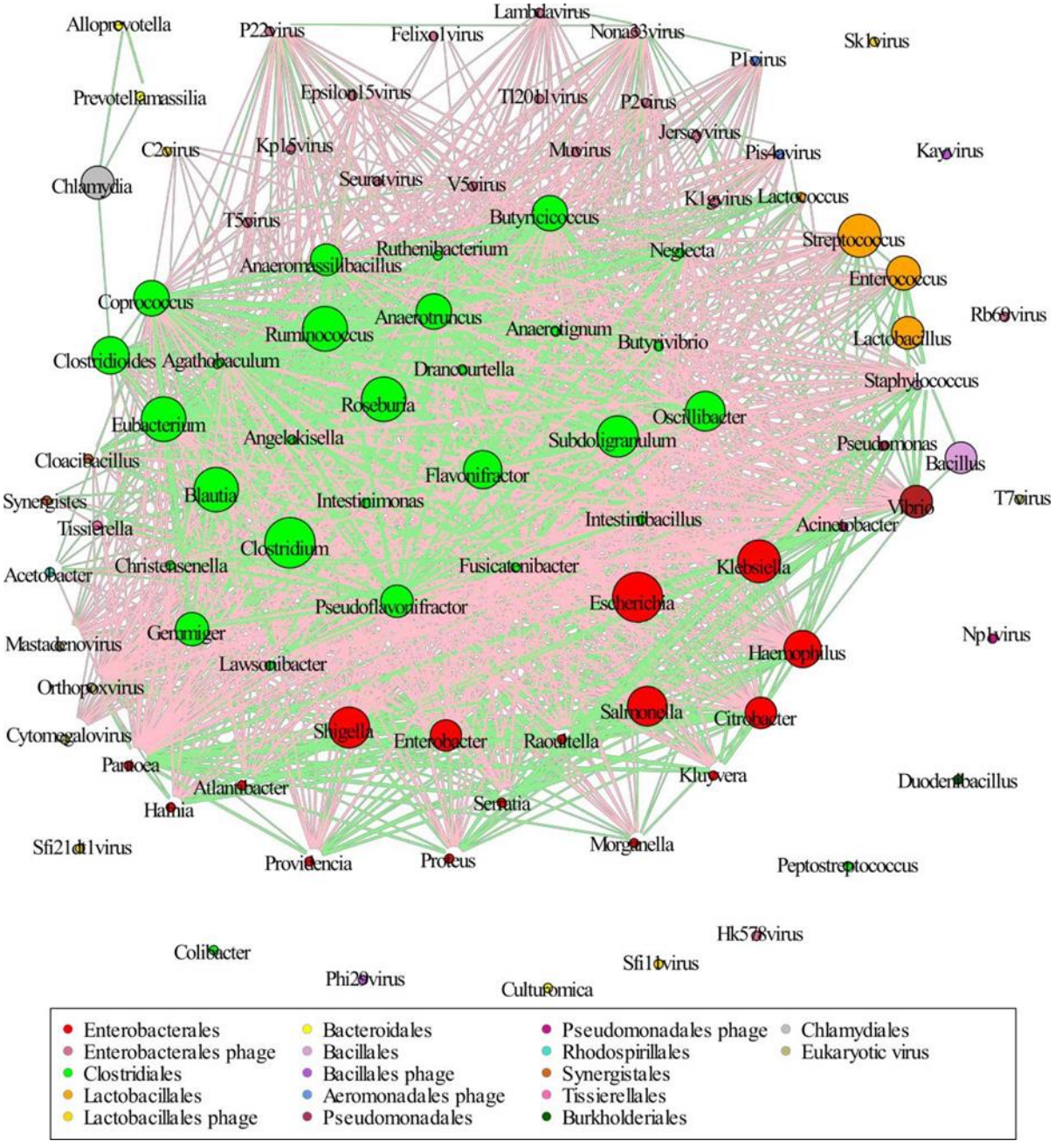
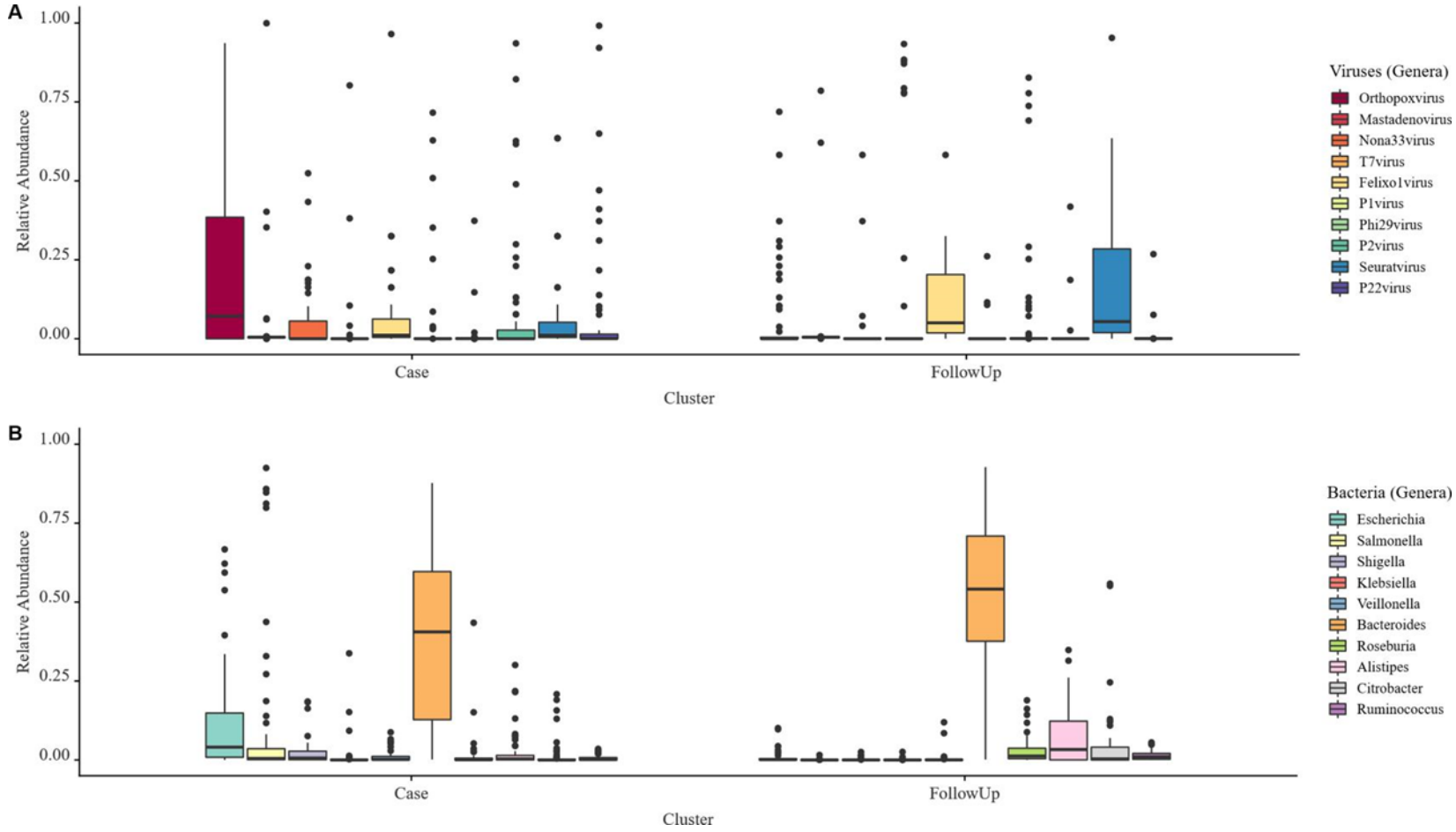


Figure 3.10. Matched microbiome from 62 Cases and their Follow-Up samples A) The top 10 differentially abundant viruses across the matched samples. B) The top 10 highest abundant bacteria across the matched samples. Both viruses and bacteria are presented at the genera taxonomical rank. The line in the box represents the median. The interquartile range (25%-75%) in the box surrounding the median. The whiskers extend from 5% to 95% of the data. Outliers are circles.



REFERENCES

REFERENCES

1. **Hall AJ, Rosenthal M, Gregoricus N, Greene SA, Ferguson J, Henao OL, Vinjé J, Lopman BA, Parashar UD, Widdowson MA**. 2011. Incidence of acute gastroenteritis and role of norovirus, Georgia, USA, 2004-2005. Emerg Infect Dis 17:1381–1388.

2. **Herkstad H, Yang S, Van Gilder TJ, Vugia D, Hadler J, Blake P, Deneen V, Shiferaw B, Angulo FJ**. 2002. A population-based estimate of the burden of diarrhoeal illness in the United States: FoodNet, 1996-7. Epidemiol Infect 129:9–17.

3. **Lupp C, Robertson ML, Wickham ME, Sekirov I, Champion OL, Gaynor EC, Finlay BB**. 2007. Host-Mediated Inflammation Disrupts the Intestinal Microbiota and Promotes the Overgrowth of Enterobacteriaceae. Cell Host Microbe 2:119–129.

4. **Schwille-Kiuntke J, Enck P, Zendler C, Krieg M, Polster A V., Klosterhalfen S, Autenrieth IB, Zipfel S, Frick JS**. 2011. Postinfectious irritable bowel syndrome: Follow-up of a patient cohort of confirmed cases of bacterial infection with Salmonella or Campylobacter. Neurogastroenterol Motil.

5. **Thabane M, Simunovic M, Akhtar-Danesh N, Garg AX, Clark WF, Collins SM, Salvadori M, Marshall JK**. 2010. An outbreak of acute bacterial gastroenteritis is associated with an increased incidence of irritable bowel syndrome in children. Am J Gastroenterol.

6. **Rodríguez LAG, Ruigómez A, Panés J**. 2006. Acute Gastroenteritis Is Followed by an Increased Risk of Inflammatory Bowel Disease. Gastroenterology 130:1588–1594.

7. **Singh P, Teal TK, Marsh TL, Tiedje JM, Mosci R, Jernigan K, Zell A, Newton DW, Salimnia H, Lephart P, Sundin D, Khalife W, Britton RA, Rudrik JT, Manning SD**. 2015. Intestinal microbial communities associated with acute enteric infections and disease recovery. Microbiome 3:45.

8. **Castaño-Rodríguez N, Underwood AP, Merif J, Riordan SM, Rawlinson WD, Mitchell HM, Kaakoush NO**. 2018. Gut Microbiome Analysis Identifies Potential Etiological Factors in Acute Gastroenteritis. Infect Immun 86:e00060-18.

9. **Singh P, Teal TK, Marsh TL, Tiedje JM, Mosci R, Jernigan K, Zell A, Newton DW, Salimnia H, Lephart P, Sundin D, Khalife W, Britton RA, Rudrik JT, Manning SD**. 2015. Intestinal microbial communities associated with acute enteric infections and disease recovery. Microbiome 3:45.

10. **Package T, Champely AS, Champely S**. 2009. Package ‘ pwr .’ October 1–21.

11. **Ihaka R, Gentleman R**. 1996. R: A Language for Data Analysis and Graphics. J Comput Graph Stat.

12. **Cohen J**. 1988. Statistical power analysis for the behavioral sciences. Stat Power Anal

Behav Sci.

13. **Bolger a. M, Lohse M, Usadel B**. 2014. Trimmomatic: A flexible read trimming tool for Illumina NGS data. Bioinformatics 30:2114–2120.

14. **Andrews S**. 2010. FastQC: A quality control tool for high throughput sequence data. babraham Bioinforma http://www.bioinformatics.babraham.ac.uk/projects/.

15. **Norman JM, Handley SA, Baldridge MT, Droit L, Liu CY, Keller BC, Kambal A, Monaco CL, Zhao G, Fleshner P, Stappenbeck TS, McGovern DPB, Keshavarzian A, Mutlu EA, Sauk J, Gevers D, Xavier RJ, Wang D, Parkes M, Virgin HW, Norman JM, Handley SA, Baldridge MT, Droit L, Liu CY, Keller BC**. 2015. Disease-Specific Alterations in the Enteric Virome in Inflammatory Bowel Disease. Cell 160:447–460.

16. **Cottingham RW**. 2015. The DOE systems biology knowledgebase (KBase).

17. **Langmead B, Salzberg SL**. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–9.

18. **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R**. 2009. The Sequence Alignment / Map (SAM) Format and SAMtools 1000 Genome Project Data Processing Subgroup. Bioinformatics 25:1–2.

19. **Menzel P, Ng KL, Krogh A**. 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun 7:11257.

20. **Agarwala R, Barrett T, Beck J, Benson DA, Bollin C, Bolton E, Bourexis D, Brister JR, Bryant SH, Canese K, Cavanaugh M, Charowhas C, Clark K, Dondoshansky I, Feolo M, Fitzpatrick L, Funk K, Geer LY, Gorelenkov V, Graeff A, Hlavina W, Holmes B, Johnson M, Kattman B, Khotomlianski V, Kimchi A, Kimelman M, Kimura M, Kitts P, Klimke W, Kotliarov A, Krasnov S, Kuznetsov A, Landrum MJ, Landsman D, Lathrop S, Lee JM, Leubsdorf C, Lu Z, Madden TL, Marchler-Bauer A, Malheiro A, Meric P, Karsch-Mizrachi I, Mnev A, Murphy T, Orris R, Ostell J, O'Sullivan C, Palanigobu V, Panchenko AR, Phan L, Pierov B, Pruitt KD, Rodarmer K, Sayers EW, Schneider V, Schoch CL, Schuler GD, Sherry ST, Siyan K, Soboleva A, Soussov V, Starchenko G, Tatusova TA, Thibaud-Nissen F, Todorov K, Trawick BW, Vakatov D, Ward M, Yaschenko E, Zasypkin A, Zbicz K**. 2018. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res.

21. **Rossum G Van**. 2018. The Python / C API. Python Softw Found.

22. **Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto JM, Bertalan M, Borruel N, Casellas F, Fernandez L, Gautier L, Hansen T, Hattori M, Hayashi T, Kleerebezem M, Kurokawa K, Leclerc M, Levenez F, Manichanh C, Nielsen HB, Nielsen T, Pons N, Poulain J, Qin J, Sicheritz-Ponten T, Tims S, Torrents D, Ugarte E, Zoetendal EG, Wang J, Guarner F, Pedersen O, De Vos WM, Brunak S, Doré J, Weissenbach J, Ehrlich SD, Bork P.** 2011. Enterotypes of the human gut microbiome. Nature.

23. **Nurk S, Meleshko D, Korobeynikov A, Pevzner PA**. 2017. metaSPAdes: a new versatile metagenomic assembler. Genome Res 27:824–834.

24. **Sanders HL**. 1968. Marine Benthic Diversity: A Comparative Study. Am Nat.

25. **Hurlbert SH**. 1971. The Nonconcept of Species Diversity: A Critique and Alternative Parameters. Ecology.

26. **Heck KL, van Belle G, Simberloff D**. 1975. Explicit Calculation of the Rarefaction Diversity Measurement and the Determination of Sufficient Sample Size. Ecology.

27. **Jari Oksanen, F. Guillaume Blanchet, Michael Friendly RK, Pierre Legendre, Dan McGlinn, Peter R. Minchin RBO, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens ES, Helene Wagne**. 2017. Package 'vegan' | Community Ecology Package 1–292.

28. **Rodriguez-R LM, Gunturu S, Tiedje JM, Cole JR, Konstantinidis KT**. 2018. Nonpareil 3: Fast Estimation of Metagenomic Coverage and Sequence Diversity. mSystems.

29. **Rohart F, Gautier B, Singh A, Lê Cao KA**. 2017. mixOmics: An R package for 'omics feature selection and multiple data integration. PLoS Comput Biol.

30. **Palarea-Albaladejo J, Martín-Fernández JA**. 2015. ZCompositions - R package for multivariate imputation of left-censored data under a compositional approach. Chemom Intell Lab Syst.

31. **Pärnänen K, Karkman A, Hultman J, Lyra C, Bengtsson-Palme J, Larsson DGJ, Rautava S, Isolauri E, Salminen S, Kumar H, Satokari R**, Virta M. 2018. Maternal gut and breast milk microbiota affect infant gut antibiotic resistome and mobile genetic elements. Nat Commun.

32. **Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ**. 2017. Microbiome datasets are compositional: And this is not optional. Front Microbiol.

33. **Aitchison J**. 1982. The Statistical Analysis of Compositional Data. J R Stat Soc Ser B.

34. **Mandal S, Treuren W Van, White RA, Eggesbø M, Knight R, Peddada SD**. 2015. Analysis of composition of microbiomes: a novel method for studying microbial composition 1:1–7.

35. **Kurtz ZD, Müller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA**. 2015. Sparse and Compositionally Robust Inference of Microbial Ecological Networks. PLoS Comput Biol.

36. **Friedman J, Alm EJ**. 2012. Inferring Correlation Networks from Genomic Survey Data. PLoS Comput Biol.

37. **Centers for Disease Control and Prevention**. 2015. Introduction to Epi Info ™ 7 Using Epi Info ™. Epi Info.

38. **Hosmer DW, Lemeshow S**. 1980. Goodness of fit tests for the multiple logistic regression model. Commun Stat - Theory Methods.

39. **David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Ling A V., Devlin AS, Varma Y, Fischbach MA, Biddinger SB, Dutton RJ, Turnbaugh PJ**. 2014. Diet rapidly and reproducibly alters the human gut microbiome. Nature.

40. **Panda S, El Khader I, Casellas F, López Vivancos J, García Cors M, Santiago A, Cuenca S, Guarner F, Manichanh C**. 2014. Short-term effect of antibiotics on human gut microbiota. PLoS One.

41. **Fettweis JM, Serrano MG, Brooks JP, Edwards DJ, Girerd PH, Parikh HI, Huang B, Arodz TJ, Edupuganti L, Glascock AL, Xu J, Jimenez NR, Vivadelli SC, Fong SS, Sheth NU, Jean S, Lee V, Bokhari YA, Lara AM, Mistry SD, Duckworth RA, Bradley SP, Koparde VN, Orenda XV, Milton SH, Rozycki SK, Matveyev A V., Wright ML, Huzurbazar S V., Jackson EM, Smirnova E, Korlach J, Tsai YC, Dickinson MR, Brooks JL, Drake JI, Chaffin DO, Sexton AL, Gravett MG, Rubens CE, Wijesooriya NR, Hendricks-Muñoz KD, Jefferson KK, Strauss JF, Buck GA**. 2019. The vaginal microbiome and preterm birth. Nat Med 25:1012–1021.

42. **Zhou W, Sailani MR, Contrepois K, Zhou Y, Ahadi S, Leopold SR, Zhang MJ, Rao V, Avina M, Mishra T, Johnson J, Lee-McMullen B, Chen S, Metwally AA, Tran TDB, Nguyen H, Zhou X, Albright B, Hong BY, Petersen L, Bautista E, Hanson B, Chen L, Spakowicz D, Bahmani A, Salins D, Leopold B, Ashland M, Dagan-Rosenfeld O, Rego S, Limcaoco P, Colbert E, Allister C, Perelman D, Craig C, Wei E, Chaib H, Hornburg D, Dunn J, Liang L, Rose SMSF, Kukurba K, Piening B, Rost H, Tse D, McLaughlin T, Sodergren E, Weinstock GM, Snyder M**. 2019. Longitudinal multi-omics of host–microbe dynamics in prediabetes. Nature 569:663–671.

43. **Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon TW, Andrews E, Ajami NJ, Bonham KS, Brislawn CJ, Casero D, Courtney H, Gonzalez A, Graeber TG, Hall AB, Lake K, Landers CJ, Mallick H, Plichta DR, Prasad M, Rahnavard G, Sauk J, Shungin D, Vázquez-Baeza Y, White RA, Braun J, Denson LA, Jansson JK, Knight R, Kugathasan S, McGovern DPB, Petrosino JF, Stappenbeck TS, Winter HS, Clish CB, Franzosa EA, Vlamakis H, Xavier RJ, Huttenhower C**. 2019. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature 569:655–662.

44. **Santisteban MM, Qi Y, Zubcevic J, Kim S, Yang T, Shenoy V, Cole-Jeffrey CT, Lobaton GO, Stewart DC, Rubiano A, Simmons CS, Garcia-Pereira F, Johnson RD, Pepine CJ, Raizada MK**. 2017. Hypertension-Linked Pathophysiological Alterations in the Gut. Circ Res.

45. **Gomez-Arango LF, Barrett HL, McIntyre HD, Callaway LK, Morrison M, Dekker Nitert M**. 2016. Increased Systolic and Diastolic Blood Pressure is Associated with Altered Gut Microbiota Composition and Butyrate Production in Early Pregnancy. Hypertension.

46. **Hiippala K, Kainulainen V, Kalliomäki M, Arkkila P, Satokari R**. 2016. Mucosal

prevalence and interactions with the epithelium indicate commensalism of Sutterella spp. Front Microbiol.

47. **Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, Rohwer F, Gordon JI**. 2010. Viruses in the faecal microbiota of monozygotic twins and their mothers. Nature 466:334–338.

48. **Reyes A, Blanton L V., Cao S, Zhao G, Manary M, Trehan I, Smith MI, Wang D, Virgin HW, Rohwer F, Gordon JI**. 2015. Gut DNA viromes of Malawian twins discordant for severe acute malnutrition. Proc Natl Acad Sci U S A 112:11941–11946.

49. **Lim ES, Zhou Y, Zhao G, Bauer IK, Droit L, Ndao IM, Warner BB, Tarr PI, Wang D, Holtz LR**. 2015. Early life dynamics of the human gut virome and bacterial microbiome in infants. Nat Med 21:1228–1234.

50. **Broecker F, Russo G, Klumpp J, Moelling K**. 2016. Stable core virome despite variable microbiome after fecal transfer. Gut Microbes 8:1–7.

51. **Norman JM, Handley SA, Baldridge MT, Droit L, Liu CY, Keller BC, Kambal A, Monaco CL, Zhao G, Fleshner P, Stappenbeck TS, McGovern DPB, Keshavarzian A, Mutlu EA, Sauk J, Gevers D, Xavier RJ, Wang D, Parkes M, Virgin HW**. 2015. Disease-specific alterations in the enteric virome in inflammatory bowel disease. Cell 160:447–460.

52. **Monaco CL, Gootenberg DB, Zhao G, Handley SA, Ghebremichael MS, Lim ES, Lankowski A, Baldridge MT, Wilen CB, Flagg M, Norman JM, Keller BC, Luvano JM, Wang D, Boum Y, Martin JN, Hunt PW, Bangsberg DR, Siedner MJ, Kwon DS, Virgin HW**. 2016. Altered Virome and Bacterial Microbiome in Human Immunodeficiency Virus-Associated Acquired Immunodeficiency Syndrome. Cell Host Microbe 19:311–322.

53. **Arumugam M, Raes J, Pelletier E, Paslier D Le, Batto J, Bertalan M, Borruel N, Casellas F, Costea PI, Hildebrand F, Manimozhiyan A, Bäckhed F, Blaser MJ, Bushman FD, De Vos WM, Ehrlich SD, Fraser CM, Hattori M, Huttenhower C, Jeffery IB, Knights D, Lewis JD, Ley RE, Ochman H, O'Toole PW, Quince C, Relman DA, Shanahan F, Sunagawa S, Wang J, Weinstock GM, Wu GD, Zeller G, Zhao L, Raes J, Knight R, Bork P, Gorvitovskaia A, Holmes SP, Huse SM**. 2013. Enterotypes in the landscape of gut microbial community composition. Nature 3:1–12.

54. **Chen S-Y, Tsai C-N, Lee Y-S, Lin C-Y, Huang K-Y, Chao H-C, Lai M-W, Chiu C-H**. 2017. Intestinal microbiome in children with severe and complicated acute viral gastroenteritis. Sci Rep 7:46130.

55. **Stecher B, Robbiani R, Walker AW, Westendorf AM, Barthel M, Kremer M, Chaffron S, Macpherson AJ, Buer J, Parkhill J, Dougan G, Von Mering C, Hardt WD**. 2007. Salmonella enterica serovar typhimurium exploits inflammation to compete with the intestinal microbiota. PLoS Biol.

56. **Small CL, Xing L, McPhee JB, Law HT, Coombes BK**. 2016. Acute Infectious

Gastroenteritis Potentiates a Crohn's Disease Pathobiont to Fuel Ongoing Inflammation in the Post-Infectious Period. PLoS Pathog 12:1–20.

57. **Larsen JM, Steen-Jensen DB, Laursen JM, Søndergaard JN, Musavian HS, Butt TM, Brix S**. 2012. Divergent pro-inflammatory profile of human dendritic cells in response to commensal and pathogenic bacteria associated with the airway microbiota. PLoS One.

58. **Ederveen THA, Ferwerda G, Ahout IM, Vissers M, de Groot R, Boekhorst J, Timmerman HM, Huynen MA, van Hijum SAFT, de Jonge MI**. 2018. Haemophilus is overrepresented in the nasopharynx of infants hospitalized with RSV infection and associated with increased viral load and enhanced mucosal CXCL8 responses. Microbiome.

59. **Chen J, Chia N, Kalari KR, Yao JZ, Novotna M, Soldan MMP, Luckey DH, Marietta E V., Jeraldo PR, Chen X, Weinshenker BG, Rodriguez M, Kantarci OH, Nelson H, Murray JA, Mangalam AK**. 2016. Multiple sclerosis patients have a distinct gut microbiota compared to healthy controls. Sci Rep.

60. **Zhang X, Zhang D, Jia H, Feng Q, Wang D, Liang D, Wu X, Li J, Tang L, Li Y, Lan Z, Chen B, Li Y, Zhong H, Xie H, Jie Z, Chen W, Tang S, Xu X, Wang X, Cai X, Liu S, Xia Y, Li J, Qiao X, Al-Aama JY, Chen H, Wang L, Wu QJ, Zhang F, Zheng W, Li Y, Zhang M, Luo G, Xue W, Xiao L, Li J, Chen W, Xu X, Yin Y, Yang H, Wang J, Kristiansen K, Liu L, Li T, Huang Q, Li Y, Wang J**. 2015. The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. Nat Med.

61. **Kasai C, Sugimoto K, Moritani I, Tanaka J, Oya Y, Inoue H, Tameda M, Shiraki K, Ito M, Takei Y, Takase K**. 2016. Comparison of human gut microbiota in control subjects and patients with colorectal carcinoma in adenoma: Terminal restriction fragment length polymorphism and next-generation sequencing analyses. Oncol Rep.

62. **Chen S, Tsai C, Lee Y, Lin C, Huang K**. 2017. Intestinal microbiome in children with severe and complicated acute viral gastroenteritis. Nat Publ Gr 1–7.

63. **Delétoile A, Decré D, Courant S, Passet V, Audo J, Grimont P, Arlet G, Brisse S**. 2009. Phylogeny and identification of Pantoea species and typing of Pantoea agglomerans strains by multilocus gene sequencing. J Clin Microbiol.

64. **Cekanaviciute E, Yoo BB, Runia TF, Debelius JW, Singh S, Nelson CA, Kanner R, Bencosme Y, Lee YK, Hauser SL, Crabtree-Hartman E, Sand IK, Gacias M, Zhu Y, Casaccia P, Cree BAC, Knight R, Mazmanian SK, Baranzini SE**. 2017. Gut bacteria from multiple sclerosis patients modulate human T cells and exacerbate symptoms in mouse models. Proc Natl Acad Sci U S A.

65. **Ness IF, Diep DB, Ike Y**. 2014. Enterococcal Bacteriocins and Antimicrobial Proteins that Contribute to Niche ControlEnterococci: From Commensals to Leading Causes of Drug Resistant Infection.

66. **Duncan SH, Hold GL, Barcenilla A, Stewart CS, Flint HJ**. 2002. Roseburia intestinalis sp. nov., a novel saccharolytic, butyrate-producing bacterium from human faeces. Int J Syst Evol Microbiol.

67. **Segain JP, Galmiche JP, Raingeard De La Blétière D, Bourreille A, Leray V, Gervois N, Rosales C, Ferrier L, Bonnet C, Blottière HM**. 2000. Butyrate inhibits inflammatory responses through NFκB inhibition: Implications for Crohn's disease. Gut.

68. **van Beek AA, Sovran B, Hugenholtz F, Meijer B, Hoogerland JA, Mihailova V, van der Ploeg C, Belzer C, Boekschoten M V., Hoeijmakers JHJ, Vermeij WP, de Vos P, Wells JM, Leenen PJM, Nicoletti C, Hendriks RW, Savelkoul HFJ**. 2016. Supplementation with lactobacillus plantarum wcfs1 prevents decline of mucus barrier in colon of accelerated aging Ercc1-/Δ7 mice. Front Immunol.

69. **Saint-Cyr MJ, Haddad N, Taminiau B, Poezevara T, Quesne S, Amelot M, Daube G, Chemaly M, Dousset X, Guyard-Nicodème M**. 2017. Use of the potential probiotic strain Lactobacillus salivarius SMXD51 to control Campylobacter jejuni in broilers. Int J Food Microbiol.

70. **Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, Beaumont M, Van Treuren W, Knight R, Bell JT, Spector TD, Clark AG, Ley RE**. 2014. Human genetics shape the gut microbiome. Cell 159:789–799.

71. **Gogokhia L, Buhrke K, Bell R, Hoffman B, Brown DG, Hanke-Gogokhia C, Ajami NJ, Wong MC, Ghazaryan A, Valentine JF, Porter N, Martens E, O'Connell R, Jacob V, Scherl E, Crawford C, Stephens WZ, Casjens SR, Longman RS, Round JL**. 2019. Expansion of Bacteriophages Is Linked to Aggravated Intestinal Inflammation and Colitis. Cell Host Microbe 25:285-299.e8.

72. **Reyes A, Wu M, McNulty NP, Rohwer FL, Gordon JI**. 2013. Gnotobiotic mouse model of phage-bacterial host dynamics in the human gut. Proc Natl Acad Sci U S A 110:20236–20241.

73. **Winter C, Bouvier T, Weinbauer MG, Thingstad TF**. 2010. Trade-offs between competition and defense specialists among unicellular planktonic organisms: the "killing the winner" hypothesis revisited. Microbiol Mol Biol Rev 74:42–57.

74. **Barr JJ, Auro R, Furlan M, Whiteson KL, Erb ML, Pogliano J, Stotland A, Wolkowicz R, Cutting AS, Doran KS, Salamon P, Youle M, Rohwer F**. 2013. Bacteriophage adhering to mucus provide a non-host-derived immunity. Proc Natl Acad Sci U S A 110:10771–6.

75. **Duerkop B a, Clements C V, Rollins D, Rodrigues JLM, Hooper L V**. 2012. A composite bacteriophage alters colonization by an intestinal commensal bacterium. Proc Natl Acad Sci U S A 109:17621–6.

76. **Wang J, Zheng J, Shi W, Du N, Xu X, Zhang Y, Ji P, Zhang F, Jia Z, Wang Y, Zheng Z, Zhang H, Zhao F**. 2018. Dysbiosis of maternal and neonatal microbiota associated with gestational diabetes mellitus. Gut.

77. **Alcami A**. 2003. Viral mimicry of cytokines, chemokines and their receptors. Nat Rev Immunol.

78. **Finlay BB, McFadden G**. 2006. Anti-immunology: Evasion of the host immune system by bacterial and viral pathogens. Cell.

79. **De Cárcer DA, Hernáez B, Rastrojo A, Alcamí A**. 2017. Infection with diverse immune-modulating poxviruses elicits different compositional shifts in the mouse gut microbiome. PLoS One 12:1–9.

80. **Kernbauer E, Ding Y, Cadwell K**. 2014. An enteric virus can replace the beneficial function of commensal bacteria. Nature 516:94–98.

81. **Yang JY, Kim MS, Kim E, Cheon JH, Lee YS, Kim Y, Lee SH, Seo SU, Shin SH, Choi SS, Kim B, Chang SY, Ko HJ, Bae JW, Kweon MN**. 2016. Enteric Viruses Ameliorate Gut Inflammation via Toll-like Receptor 3 and Toll-like Receptor 7-Mediated Interferon-β Production. Immunity 44:889–900.

82. **Zhao G, Wu G, Lim ES, Droit L, Krishnamurthy S, Barouch DH, Virgin HW, Wang D**. 2017. VirusSeeker, a computational pipeline for virus discovery and virome composition analysis. Virology 503:21–30.

CHAPTER 4

ISOLATION OF BACTERIOPHAGES FROM THE HUMAN GUT THAT CAN LYSE ENTERIC PATHOGENS AND REPRESS SHIGA TOXIN PRODUCTION

ABSTRACT

Bacteriophages are viruses that infect bacteria and are found in many environments, including the human gastrointestinal (GI) tract. The role that these viruses play in human health, however, is not well understood. The goal of this study was to isolate and characterize virus-like particles (VLPs), or bacteriophage communities, from the stools of patients with enteric infections and healthy individuals, and to evaluate their impact on enteric pathogens. Bacteria-bacteriophage interactions were evaluated using spot tests to examine the ability of the isolated VLPs to lyse three commensal *Escherichia coli* strains and three common enteric pathogens (*E. coli* O157:H7, *Salmonella* Typhimurium, and *Shigella sonnei*). Notably, the isolated VLPs lysed pathogenic strains at higher frequencies (78%) when compared to commensal strains (39%). Among the viral communities, Poxviridae and Anelloviridae were more abundant in samples from patients with acute bacterial gastroenteritis, while Caudovirales predominated across all samples. Isolation of three bacteriophages for genomic sequencing and characterization identified two related lysogenic phages (PHG002 and PHG003) and one lytic bacteriophage (PHG001). Homologous sequences for this lytic bacteriophage were found in 87% of the 15 sequenced viral communities. In culture, PHG001 reduced *E. coli* O157:H7 growth by 3-fold after 3 hours but had no bactericidal effect on three commensal *E. coli* strains. Importantly, PHG001 contributed to a 16-fold reduction in the expression of Shiga toxin genes by *E. coli* O157 at 3 hours. These results suggest that bacteriophage populations residing in the gut may play an important role in pathogen control and that further characterization of these populations is warranted.

INTRODUCTION

Bacteriophages (phages) are ubiquitous viruses that infect bacteria. Phages are classified either as lytic due to their ability to lyse specific bacteria following replication within the host; or as temperate phages that remain dormant in host bacteria until induction (1). Phages have been cultured directly from human fecal samples, and the use of metagenomics has shown that phages are the most abundant double-stranded deoxyribonucleic acid (DNA) viruses in the gut (2, 3). Phages are critical for shaping the composition, diversity, and function of bacterial populations. They also exert selective pressures on bacteria, which can contribute to resistance to subsequent bacteriophage infections and alterations in the resident bacterial population (1). Additionally, bacteriophages can impact competition between species and mediate horizontal gene transfer (1, 4). Gene transfer can introduce phenotypic changes into bacterial communities, alter bacterial metabolic profiles, and impact host immune responses (5, 6).

Unlike bacterial populations in the gut, which are highly similar among related individuals (7, 8), a study of monozygotic twins and their mothers found that virus populations were unique per individual in both the types of viruses present as well as their functional gene profiles (3). Other studies of unrelated individuals also observed a high degree of interpersonal variation (5, 9) and have found that virus populations were relatively stable within individuals over time (3, 6). Nonetheless, the human gut serves as a reservoir for viruses, particularly phages that are common among individuals (9). Metagenomics data from 124 unrelated individuals, for example, revealed that 29% of the bacteriophage contigs were present in at least 10% of the individuals examined (9). Little is known, however, about whether specific bacteriophage populations are essential for either inhibiting or exacerbating acute bacterial gastroenteritis infections.

Bacteriophages have also been shown to prevent pathogen invasion of epithelial cells in mucosa cell lines through direct infection and lysis of the pathogen (10). Given this finding as well as the high degree of variation and stability among intestinal viral populations across individuals, we sought to investigate the role that intestinal bacteriophage populations play in enteric infections. It is indeed likely that variation in bacteriophage composition, abundance, and function within distinct intestinal communities can contribute to differences in susceptibility to enteric infections as well as disease severity and recovery. Hence, we isolated and sequenced virus-like particles (VLPs) from the stools of patients with enteric infections for comparison to those from healthy individuals. VLPs were examined for their ability to infect common bacterial pathogens and commensal E. coli strains, which are typically found in the GI tract and do not cause disease. A bacteriophage specific for Shiga toxin-producing E. coli (STEC) O157:H7 was isolated, sequenced, and used to evaluate host range and impact on Shiga toxin (stx) expression in vitro as well as its abundance within the sequenced viral communities. Examining the relationship between bacterial pathogens and bacteriophage populations within the gut during infection is critical to enhance understanding of the disease process and could lead to ideas for the development of new therapies and prevention methods.

MATERIALS AND METHODS

Sampling and isolation of virus communities

As described previously (11), stool samples were collected via the Michigan Department of Health and Human Services (MDHHS) from patients with bacterial gastroenteritis (cases) caused by *Campylobacter jejuni,* non-typhoidal *Salmonella* spp., *Shigella* spp., or Shiga-toxin producing *Escherichia coli* (STEC). Contract tracing was used to identify and enroll healthy family members of the patients (controls). For this study, 18 samples from patients (n=14) and healthy family members (n=4) were selected for isolation and characterization of the stool-derived viral communities. The 14 cases included patients with acute infections caused by *C. jejuni* (n=4), *Salmonella* spp. (n=4), *Shigella* spp. (n=4), and STEC (n=2).

Polyethylene glycol (PEG) precipitation was used to recover the virus-like particles (VLPs) from each stool sample. In brief, stools were centrifuged at 4,000 RPM for 10 minutes to pellet debris, and the supernatant was collected for PEG precipitation as described (12). PEG (molecular weight = 8000) with NaCl (2.5% w/v) was added to 1/6 final volume of the supernatant and the mixture was inverted twice and stored at 4^{o} C. The sample was centrifuged at 11,000 g for 10 minutes and the pellet was resuspended in 15 ml of bacteriophage buffer containing Tris (10 mM), pH 7.5, MgSO4 (10 mM), NaCl (68 mM), and CaCl2 (1 mM), filter-sterilized (0.22 µm) and stored at 4^{o} C. All stool samples obtained in this study were previously approved for collection and use by the Institutional Review Boards at Michigan State University (MSU; Lansing, MI, USA; IRB #10-736SM) and the MDHHS (842-PHALAB).

Spot testing, and quantification of viruses by plaque assays

Spot tests were first performed with three pathogens and three commensals to classify the host range of the VLPs recovered from all 18 samples. The three pathogens, which were previously

isolated from patients with enteric infections, included: *Shigella sonnei* (TW16372), *Salmonella enterica* serovar Typhimurium (TW16390), and STEC *E. coli* O157:H7 (TW14359 (13)). Three commensal *E. coli* strains (TW17000, TW17041, and TW17368), recovered from the stools of healthy individuals, were evaluated for comparison. Each commensal *E. coli* strain was confirmed to lack genes encoding common STEC virulence factors including *stx* (Shiga toxin) and *eae* (intimin) by polymerase chain reaction (PCR) as described previously (13). The absence of other virulence-associated genes such as *escV*, *bfpA*, *ipaH*, *estla*, and *elt*, which are also common among pathogenic *E.coli* (14), was confirmed by genome sequencing.

Spot testing was performed by growing each of the six bacterial strains in Luria Broth (Sigma-Aldrich, St. Louis, MO) to exponential phase at an optical density (OD_{600}) of 0.2; 300 µl of bacterial cells were added to 3 ml soft agar (0.5%), mixed by inversion, and poured onto LB agar. After the agar solidified, 10 µl of supernatant containing the stool derived VLPs from the 14 patients, and four healthy individuals were spotted on to each of the six bacterial lawns. Plates were incubated at room temperature for 20 minutes, followed by overnight incubation at 37°C. Plates were evaluated for any bacterial lysis. Clearance at the site where VLPs were added to the bacteria lawn was classified as lysis, which is indicative of the presence of bacteriophage within the VLP community that could inhibit bacterial growth. The Chi-square test was used to detect differences in lysis by bacterial type, and Odds Ratios (OR) and their corresponding 95% confidence intervals (CI) were calculated in Epi Info™ v.7 (15).

The 18 viral communities were subjected to plaque assays using the double-layer method in order to quantify the abundance of VLPs infecting the six bacterial strains (16). Briefly, VLP stocks were serially diluted 10-fold in bacteriophage buffer, and 100 µl added to 300 µl of exponentially growing (OD_{600} = 0.2) bacteria. VLPs were allowed to adsorb to the bacterial host

by incubating at room temperature for 15 minutes. The VLP-bacteria co-culture was then added to 3 ml of soft agar (0.5%), mixed gently, plated on LB agar, and incubated overnight at 37° C. Bacteriophages form plaques on soft agar when grown together with specific bacterial hosts. These plaques, which often represent single bacteriophages that can infect and replicate in the host bacteria, can be quantified as plaque-forming units (PFUs)/ml for each bacterial strain. All assays were repeated in triplicate.

Metagenomics of virus communities

All 18 stool-derived VLP communities were sequenced, though only 15 samples yielded good-quality reads (PHRED > 30) for inclusion in the analysis. In brief, DNA libraries were prepared using the PicoPLEX kit (Rubicon Genomics, Ann Arbor, MI, USA). After quality control checking and quantitation, this library pool was loaded onto an Illumina MiSeq V2 flow cell and sequenced using a standard 500 cycle reagent kit in a 2x250bp paired-end format (Illumina, San Diego, CA, USA) at the Michigan State University Research Technology Support Facility (RTSF). Base calling was performed with Illumina Real-Time Analysis (RTA) v1.18.54, and the RTA output data were demultiplexed and converted to FastQ format with Illumina Bcl2fastq v1.8.4. Coverage for each metagenome was estimated using Nonpariel3 (17).

Reads were quality trimmed with Trimmomatic (18), and human reads were removed with Bowtie2 (19). Kaiju (20) was used to annotate quality-controlled reads to the non-redundant protein database in the National Center for Biotechnology Information (NCBI), and community viral profiles were generated by filtering for the top 1% of reported taxa. Zero counts were replaced with multiplicative simple replacement using the zCompositions package (21) in R (22). Profiles were total sum scaled, log-transformed, and visualized using ggplot (23). Diversity, richness, and

evenness were examined using vegan (24). All unassembled metagenome sequences were submitted to the NCBI) Sequence Read Archive.

Bacteriophage isolation and propagation

Individual plaques capable of lysing the *E. coli* O157:H7 (TW14359) strain in the plaque assays were picked for isolation and further characterization. In brief, plaques with unique morphologies were picked using a sterile 10 µl pipette tip, mixed with 200 µl of bacteriophage buffer, filtered with a 0.22 µm filter, and stored at 4°C. Bacteriophage stocks were created for TW14359 infections grown to an OD_{600}= 0.2. Co-cultures were incubated aerobically at 37°C overnight, and 250 ml were aliquoted into separate 50 ml tubes and centrifuged for 10 minutes at 4,000 RPM. The supernatant was filtered using a 0.22 µm filter, pooled and stored at 4°C for bacteriophage quantification, scanning electron microscopy (SEM), genome sequencing, and host range testing. Plaque assays were performed to quantify bacteriophage concentrations.

Sequencing of bacteria and bacteriophage genomes

The six bacterial strains used for the spot testing and host range analysis of VLPs were sequenced as were three isolated bacteriophages (PHG001, PHG002, and PHG003) recovered from the *E. coli* O157:H7 (TW14539) infections. Bacterial DNA was extracted using the Qiagen DNA Extraction kit (Qiagen Sciences, MD, USA), while bacteriophage DNA was isolated using the Phage DNA Isolation Kit (Norgen Biotek, Thorold, ON, Canada) per the manufacturer's guidelines. A single pool of DNA libraries was prepared separately for the bacteriophage and bacterial samples using the Illumina Nextera Library Preparation Kit. Quality control and quantification of each library was performed using the following assays: Qubit dsDNA HS (Thermo Fisher Scientific, Waltham, MA, USA), Caliper LabChipGX HS DNA, (Caliper Life Sciences, Hopkinton, MA, USA) and Kapa Illumina Library Quantification qPCR (Kapa

Biosystems, Inc, Wilmington, MA, USA). Libraries were loaded on to a MiSeq Nano v2 flow cell and sequenced using a 500 cycle (PE250) v2 reagent kit (Illumina) at the MSU RTSF; the bacterial and bacteriophage DNA samples were sequenced separately. Similar to the metagenomics analysis, base-calling was performed with Illumina RTA v1.18.54, and FASTQ files were created on the demultiplexed output of RTA. Raw reads were trimmed with Trimmomatic (18) to remove ambiguous reads, low-quality reads, and adaptors. The quality of trimming was assessed with FastQC (25), while assemblies were performed using SPADES 3.6 (26). The reads were mapped using Bowtie2 (19)**,** and genomes were annotated with Prokka (27). Functional annotation was performed using the Rapid Annotations using Subsystems Technology (RAST) server (28), and blastn (29) was used to find similar genomes in the NCBI genomic RefSeq database with an e-value set at $<10^{-5}$ (30). Assembled bacteriophage genomes were uploaded to the PHASTER server (31), an optimized version of PHAST that utilizes blast to identify and annotate prophage genomes. The output from PHASTER was downloaded and combined with the RAST and Prokka annotations that were performed on each genome.

Prophages found in other bacterial genomes were identified with PHASTER and the genomes were downloaded. Related prophage genomes were aligned globally with Progressive MAUVE (32) to identify homologous regions. The genomes of PHG001, PHG002, and PHG003 were uploaded to the ViPtree server for phylogenetic determination. In brief, ViPtree uses tBLASTx for phylogenetic analyses (33). A proteomic tree can be constructed based on the bacteriophage genome similarities compared to the viral host-db database, which allows for dendrogram based on established viral taxonomy, as demonstrated previously (33). Finally, the three bacteriophage genomes were blasted against the 15 VLP metagenomes to determine how frequently these and related phages were found within the viral communities examined in this

study. Sequencing reads were assembled using metaspades (34) from each virome, and assemblies were aligned with tblastx (e-value $<10^{-10}$) (29) to the assembled genomes. Scripts were developed in Python and are available on GitHun/BrianNo.

To more comprehensively test the host range of lytic bacteriophage PHG001, spot testing was performed using 29 *E. coli* strains from the *E. coli* Reference Collection (ECOR) (Table 4.1.). This genetically diverse group of strains was initially recovered from humans without infections and comprised strains with multiple O-antigen types (35); the ECOR strains were classified as commensal *E. coli* in our analysis. Spot testing with PHG001 was also performed with 37 additional STEC strains representing serogroups O157, O103, O111, O45, and O26, which are commonly associated with clinical infections in the United States. (36). These STEC isolates were recovered from patients in Michigan, as described in our prior study (37), and were classified as pathogens. All isolates were obtained from the STEC Center at MSU (www.shigatox.net).

Bacteriophage infection of *E. coli* O157:H7 and burst size calculation

PHG001 bacteriophage was used to infect *E. coli* O157:H7 strain TW14359 (OD_{600} of 0.2) at a multiplicity of infection (MOI) of 1. The bacteriophage titer was calculated every 20 minutes for the first 2 hours and again at 3, 4, 5, and 24 hours. Plaque forming units (PFUs) were quantified by plaque assay, and the burst size was calculated as described previously (38). Briefly, PFUs were plotted over time, and the latency period and burst size were determined. Latency was defined as the initial period of no change in PFU growth, while the burst size was determined by examining the time points before and after the burst. Assays were performed in triplicate, and the mean and standard deviation were calculated in R (22).

PHG001 impact on *E. coli* O157:H7 survival and toxin production

The level of *E. coli* O157:H7 inhibition by the PHG001 bacteriophage was compared to the level of inhibition by antibiotics. Two O157:H7 strains were used for these experiments, including the Spinach outbreak strain, TW14359 (13), which contains *stx2* and *stx2c*, and TW14313, which is positive only for *stx2*. The latter strain was evaluated for a subset of experiments, given that both strains had high levels of Stx production in a prior study (39). Ampicillin (3.8 µg/ml), mitomycin C (10 µg/ml), and PHG001 bacteriophage ($1x10^8$ PFU/ml) were examined as well as a combination of 3.8 µg/ml ampicillin or ten µg/ml mitomycin C with $1x10^8$ PFU/ml bacteriophage. Each of these treatments was added to *E. coli* O157:H7 grown to an OD_{600} of 0.2 and incubated aerobically at 37°C. Bacterial colony-forming units (CFUs) were quantified before bacteriophage challenge and 1, 2, 3, 4, 5, and 24 hours after challenge; irregular and regular colony morphologies were quantified at each time point. Experiments were performed in technical triplicate and repeated three times (n = 3). Data were log-transformed, and two standard deviations were plotted for each time point.

To evaluate *stx* expression, RNA was extracted from *E. coli* O157:H7 strain TW14359 (*stx2*, *stx2c*) cells following co-culture with PHG001 (MOI=1) and exposure to mitomycin C (10 µg/ml) using the RNAeasy Minikit (Qiagen, Germantown, MD, USA). Comparisons were made to a mock infection consisting of only bacteriophage buffer. The Turbo DNA free kit (Ambion, Foster City, CA, USA) was used to remove DNA contamination, which was confirmed by the lack of amplification of the bacterial 16S rRNA gene using PCR. The iScript Select cDNA synthesis kit (Bio-Rad, Hercules, CA, USA) was used to generate cDNA from 1 µg of RNA, while quantitative real-time PCR (qRT-PCR) was performed with the iQ SYBR Supermix kit (Bio-Rad, Hercules, CA, USA) in 15 µl reactions with 10 µM primers specific to *stx2c* as described (40). A

CFX384 Touch Real-Time PCR detection system (Bio-Rad, Hercules, CA, USA) used with one cycle of 3 minutes at 95°C and 39 cycles of the following: 95°C for 10 sec and 60°C for 30 sec. The comparative threshold cycle (Ct) method ($2^{-\Delta\Delta CT}$) was used to calculate relative gene expression levels using *gyrA* as an internal control (41). Expression was normalized to basal *stx2c* expression in the untreated O157:H7 strain and a 2-fold change or higher in gene expression were considered biologically significant.

Screening additional host backgrounds for infectivity by PHG001

Isolated bacteriophage, PHG001, was also examined for its effect on the growth of commensal *E. coli* strains, TW17000 and TW17041, and pathogenic *E. coli* strains TW18499 and TW18502 grown at an OD_{600} of 0.2 and infected with a MOI of 1. Bacterial CFUs were quantified at each time point following infection with PHG001 by plating 50 µl of bacterial cells onto LB agar and incubating at room temperature for 20 minutes followed by overnight incubation at 37°C. Bacterial colonies were classified as regular or irregular in shape and counted.

PHG001 was also used to infect *E. coli* O157:H7 strain TW14313, a previously characterized strain known for high Stx2 production (39), at two different concentrations to examine its effect on bacterial growth with increased concentration. PHG001 bacteriophage was added at $1x10^8$ PFU/ml or $1x10^9$ PFU/ml to bacterial cells grown to an OD_{600} of 0.2. Bacterial CFUs were quantified before (time point =0) and after 1, 2, 3, 4, 5, and 24 hours of growth. All bacteriophage-bacteria co-culture experiments, unless otherwise noted, were performed in triplicate and repeated three times. The mean and standard deviation were calculated for each experiment using R (22). Graphad Prism was used to visualize results.

RESULTS

Variation in the abundance of lytic virus-like particles (VLPs)

Stool derived VLPs isolated from four healthy individuals (controls) and 14 patients with enteric infections (cases) were evaluated for the ability to lyse common enteric pathogens and commensal *E. coli*. Intriguingly, the VLPs were significantly more likely to lyse the pathogens than the commensals (OR: 6.1; 95% CI: 2.60, 14.50) regardless of source. On average, 76.2% of the patient derived VLPs lysed the three pathogenic strains (n=42). VLPs isolated from patients lysed 78.6% (n=33) of *S.* Typhimurium (TW16390) challenges (n=42). Similar trends were noted with the VLPs isolated from patients and the other pathogen strains; *S. sonnei* (TW16372) was lysed 78.6% (n=33) of the time and STEC O157:H7 (TW14359) was lysed 71.4% (n=30) of the time (Table 4.1.). The results, however, were not statistically significant across patients with infections caused by different types of pathogens (Kruskal-Wallis test $p > 0.05$). By contrast, the four VLP communities isolated from healthy individuals showed growth inhibition of the three pathogens in almost all (91.7%) of the 12 infections (Table 4.1.). Healthy VLPs inhibited *S.* Typhimurium (TW16390) and STEC O157:H7 (TW14359) growth for all 12 (100%) infections, but only nine (75%) of the *S. sonnei* (TW16372) infections were inhibited. The results were not significantly different based on the pathogen tested as specific pathogens were not more likely to be inhibited than others (Kruskal-Wallis test $p > 0.05$).

Infection of the three commensal *E. coli* strains (TW17000, TW17041, TW17368) by all 18 VLP communities also resulted in variable inhibition patterns, though fewer commensal strains were inhibited when compared to the pathogenic strains. Specifically, the control derived VLPs were more likely to inhibit growth in the 12 commensal strain infections (n=10; 83.3%) compared to the 42 patient-derived VLPs infections (n=31; 73.8%), yielding an odds ratio (OR) of 5.4 (95%

confidence interval (CI): 1.3, 28.1). On average, only 31.3% of the 14 cases derived VLP communities inhibited growth in the three commensal strains. These 14-case associated VLPs prevented growth in 21.4% (n=3) of the commensal *E. coli* TW17000 infections, 42.9% (n=6) of the *E. coli* TW17041 infections and 28.6% (n=4) of the *E. coli* TW17368 infections (Table 4.1.). Among the 14 VLP communities from cases, there was no significant difference in inhibition frequencies in the three commensal *E. coli* strains based on the type of pathogen causing the patient's infection (Kruskal-Wallis test $p > 0.05$). By contrast, the four VLP communities from healthy individuals showed growth inhibition in 67.7% (n=8) of the 12 infections with all three commensal strains combined. VLPs from healthy individuals contributed to inhibition in half of the four commensal *E. coli* TW17000 and TW1704 infections, and in 100% of the four *E. coli* TW17368 infections (Figure 4.1A). The results, however, were not significantly different based on the commensal strain tested (Kruskal-Wallis test $p > 0.05$).

The VLPs also varied in abundance across samples. The average abundance of bacteriophage was $1x10^9$ PFU/ml for the three pathogen infections, which was slightly higher than the average abundance ($1x10^8$ PFU/ml) for the three commensal strains (Figure 4.1B). The difference in mean abundance between pathogens and commensals was not significant (Mann Whitney U test $p > 0.05$). Although the four VLP communities from healthy individuals had a lower average abundance of bacteriophage ($5.1x10^8$ PFU/ml) in the pathogen hosts when compared to the 14 VLPs from patients stool with acute infections ($1.3x10^9$ PFU/ml), the difference in means was also not significant (Mann Whitney U test $p > 0.05$). Overall, the highest bacteriophage titers on average were observed in the *S. sonnei* (1.7×10^9 PFU/ml) and STEC O157:H7 (1.1×10^9 PFU/ml) strains, while the broadest range of PFUs/ml (0 to 10^9) was observed in *S. sonnei*.

Coverage and annotation in metagenomes do not vary by case status

All 18 of the VLPs were submitted for sequencing; three samples (two cases, one control) did not sequence well and were not included in the analysis. A total of 1,3765,249 paired forward reads were sequenced across the 15 samples (917,683 reads per sample). A higher sequencing depth with an average of 949,232 reads was achieved in the case samples compared to the controls (791,485 reads); this difference was not significantly different (Mann Whitney U test $p > 0.05$). The average coverage as determined by Nonpareil3 (17) across all 15 viromes was 73.7%. Although cases had less coverage (69.2%) compared to controls (91.9%), this was not a significant difference (Mann Whitney U test $p > 0.05$).

Across all 15 samples, an average of 39.8% of reads fell below quality filtering parameters. Controls had a more substantial proportion of reads removed (44.5%) compared to cases (38.5%), though the difference in proportions was not significant (Mann Whitney U test $p > 0.05$). Overall, an average of 24.6% of quality-controlled reads was annotated as human. The presence of human DNA differed by case status; case samples contained 30.5% human reads compared to 1.1% in control samples, though this difference was not statistically significant (Mann Whitney U test $p > 0.05$). Kaiju annotated 17% of the reads that passed quality control checks. Cases achieved a higher annotation frequency (18.9%) compared to controls (9.2%), but the difference in frequencies was not statistically significant (Mann Whitney U test $p > 0.05$). It is possible that the low sample size (n=15), unbalanced design of the comparison groups (12 cases, 3 controls), or the wide variability in samples for each parameter tested, contributed to this finding (Table 4.2.).

Metagenomics reveals diversity within isolated virus communities

Among the 15 sequenced VLPs, the Shannon diversity index was 2.34 ± 0.81. No difference was observed in the Shannon diversity in the 12 samples from patients (2.29 ± 0.9)

compared to the three samples from healthy individuals (2.57 ± 0.2). By contrast, the richness was significantly higher in the healthy (357 ± 56) versus patient (151 ± 58) samples (Mann Whitney U test p = 0.004). The richness was 192 ± 100 among all samples combined, which is more similar to the values observed for patients given the unbalanced study design. Interestingly, the evenness was almost identical in both the patients (0.47 ± 0.18) and healthy participants (0.43 ± 0.02).

A high degree of variation was observed in the distribution and abundance of viral families across samples. The dominant families across all 15 samples were Siphoviridae, Myoviridae, Podoviridae, and Microviridae, which typically comprise bacteriophages and represented 92% of the viromes on average (Figure 4.2). Because only three viral communities from healthy individuals were available for analysis, our ability to examine differences by the source was limited. Nonetheless, we did observe an increased abundance of Siphoviridae in healthy versus patient samples comprising 78% and 50% of the virome, respectively. We also found that members of the Myoviridae family were more abundant in the patient communities (33%) compared to healthy (12%). No difference was observed among the Podoviridae and Microviridae families by case status.

Eukaryotic viruses, which belong to Poxviridae, Pithoviridae, Anelloviridae, Mimiviridae, Nimaviridae, and Phycodnaviridae, were detected in most samples even though the relative abundance of each varied. On average, the eukaryotic virus families were five times more abundant in the patient samples than the healthy samples; this difference was not statistically significant, which may be due to the small sample size. The most abundant eukaryotic virus family was Poxviridae. The average relative abundance of Poxviridae was 40% in the 12 patient samples and less than 0.0001% in the three control samples. Notably, a wide range of Poxviridae abundance was observed among the patient samples (<.0001% to 41.0%). Exclusion of the sample with the

highest proportion of Poxviridae, however, still indicated that patients had a 326-fold greater abundance of Poxviridae than healthy individuals did. Similarly, members of the Anneloviridae family were 55-fold more abundant in patients, though a wide range (<.0001% to 49.0%) was also observed and only three healthy samples were evaluated.

Diversity of bacteriophages capable of inhibiting STEC O157:H7

Three plaques with unique morphology were identified following infection of STEC O157:H7 (TW14359) by VLPs isolated from two patient stool samples. These plaques varied in morphology as well as in genomic features; two were classified as lysogenic phages and one as a lytic bacteriophage. The two lysogenic phages, PHG002 and PHG003, were highly similar to each other with an identity of 99%. Both phages were most closely related to two known phages, *Escherichia* virus pro_147 and *Escherichia* phage pro483, belonging to the Myoviridae family that use Gammaproteobacteria as hosts (Figure 4.3A). Based on tblastx alignments, PHG002 and PHG003 form a distinct cluster with a lysogenic bacteriophage found within the STEC O157:H7 host strain, TW14359. Because there are variable regions within each bacteriophage genome and some regions did not align entirely (Figure 4.3B), additional *E. coli* O157:H7 (taxid:83334) genomes were interrogated for the presence of these prophages. Notably, the use of blastn (query coverage > 99%, percent identity > 99%) identified 23 additional O157:H7 strains that possess a variant of PHG003, which is more distally related to the TW14359 bacteriophage than the PHG002 bacteriophage despite having homologous regions (Figure 4.4). Additional analyses identified a host protein exonuclease, *sbcC*, from 241 bps to 2208 bps in PHG003 that was also found in the genomes of 13 of the 23 related bacteriophage s examined. Also, a screen of the 15 sequenced viromes showed that 93% (14/15) of the viromes have sequences homologous to both PHG002 and PHG003. By contrast, the lytic bacteriophage, PHG001, has a single contiguous consensus

sequence (114,632 kbp) with coverage of 934x (Table 4.3.). Following annotation, 158 coding sequences, 187 genes, and 22 tRNAs were identified (Figure 4.5). The annotated subsystem features include 14 bacteriophage-related genes, and six metabolism-related genes; one involved in RNA metabolism, one in protein metabolism, and four in nucleoside and nucleotide metabolism. The PHG001 lytic bacteriophage is unique even though the VipTree phylogeny shows that it is related to several previously characterized bacteriophages, including *Salmonella* virus Stitch, *Escherichia* virus EPS7, *Salmonella* phage 188970_sal and *Salmonella* phage 100268_sal2 (Figure 4.6A). These related bacteriophage s were all classified as members of the Siphoviridae family and utilize Gammaproteobacteria as hosts. Based on tblastx alignments, PHG001 is most closely related to *Salmonella* virus Stitch (92% query coverage, 97.0% identity) but is distinct from both *Salmonella* phage 118970_sal2 and *Salmonella* phage 100268_sal2. Screening the 15 viral metagenomes detected the conserved regions of the PHG001 genome in 13 of the 15 (87%) samples based on tblastx alignments (e-value $<10^{-10}$).

Bacteriophage PHG001 has a broad host range

The host range of PHG001 was determined by examining its effect on the growth of 71 *E. coli* strains of various serotypes and origins as well as the *Shigella sonnei*, and *Salmonella* Typhimurium strains evaluated in the VLP analysis (Table 4.4.). Among all 73 strains, PHG001 inhibited the growth of 14 (34.1%) of the 41 pathogens and 10 of the 32 (31.3%) commensals. Strains belonging to specific *E. coli* serogroups were more commonly inhibited relative to others regardless of their source. For example, PHG001 more commonly inhibited commensal strains belonging to O6 (n=2; 100%) and O7 (n=3; 75.0%). Inhibition was also observed for one O2:nonmotile (NM) (50%), one O4:NT (50%), and three non-typeable (NT; 33.3%) commensal strains. Among the 39 pathogenic *E. coli* strains, all seven (100%) O157:H7 strains were inhibited

by PHG001 as were half (n=4) of the O26:H11 strains and two of the eight (25.0%) O103:H2 strains. No inhibition was observed for *S.* Typhimurium, the eight STEC O111:H8 strains, or the eight O45:H2 strains; however, the *S. sonnei* strain was inhibited.

It is interesting to note that the year of isolation may play a role in inhibition rates within serogroups. The four O26 strains with inhibition were recovered in 2014, whereas the four strains without inhibition were recovered in 2010. Similarly, two of the eight O103 strains with inhibition were recovered in 2010, while the four O103 isolates recovered in 2014 were not inhibited by PHG001. Although the sample sizes were small, these data suggest that isolates from specific time frames may be more similar to each other and hence, more susceptible to infection. By contrast, all seven O157:H7 strains inhibited by PHG001 were recovered in different years dating as far back as 2002, indicating that O157 strains are a primary host for this novel bacteriophage.

Bacteriophage PHG001 growth in the *E. coli* O157:H7 host

PHG001 was added to TW14359 (O.D. =0.2) and quantified to determine the bacteriophage concentration using a plaque assay. Samples were collected every 20 minutes for the first two hours following co-infection to generate a one-step growth curve to determine burst size (Figure 4.7). These samplings were followed by hourly samplings until hour five and one final sampling at hour 24. PHG001 growth shows that there is a 20-minute latency phase as there is no increase in bacteriophage concentration for this duration, and the burst size was 123 bacteriophage per hour of growth. Three rounds of infection were identified through hour 2, and PHG001 concentration increased linearly until hour five to a concentration of $1x10^7$ PFU/ml. The goal of these experiments was to assess the replication ability of the bacteriophage in the STEC O157:H7 host. We expect that the primary host will continue to grow as the MOI was low to keep the host

growing and provide a way to measure the amount of bacteriophage produced with each infection cycle.

Ampicillin and bacteriophage affect the growth of *E. coli* O157:H7

To evaluate the impact on *E. coli* O157:H7 (TW14313) growth over 24 hours, PHG001 ($1x10^8$ PFU/ml, MOI=1) was added to *E. coli* ($1x10^8$ CFU/ml). TW14313 had a 3-fold reduction at 3 hours compared to the mock culture. The inhibition of TW14313 was dose-dependent, increasing the concentration of PHG001 by 10-fold ($1x10^9$ PFU/ml, MOI=10) further reduced the growth of TW14313 5-fold by hour 4 (data not shown). Ampicillin (3.8 μg/ml) added to exponentially growing bacteria without the presence of bacteriophage, demonstrated a 2-fold reduction (99% reduction) in *E. coli* O157:H7 growth by hour 5 (Figure 4.8). At 24 hours, the ampicillin treated culture was equal to the initial bacterial concentration of $1x10^8$ CFU/ml, demonstrating that the ampicillin had little effect on bacterial growth over time. Despite an initial 3-fold (99.9%) reduction by 5 hours, the bacteria treated with bacteriophage alone ($1x10^8$ PFU/ml) fully recovered against the bacteriophage challenge, reaching the same final concentration as the control ($1x10^9$ CFU/ml) by 24 hours (data not shown).

Notably, the combination of ampicillin (3.8 μg/ml) and bacteriophage ($1x10^8$ PFU/ml) resulted in a 5-fold reduction in *E. coli* O157:H7 growth by hour two, and by five hours, there were no recoverable colonies (Detectable limit = 100 CFU/ml) (Figure 4.8). No colonies could be recovered after 24 hours as well (data not shown). Although regrowth of the bacterial population observed in the ampicillin and bacteriophage single treatments, the combination treatment showed a more rapid reduction, which prevented a rebound in bacterial growth after 24 hours. These data suggest a synergistic relationship between ampicillin and PHG001 and differential methods of inhibition. Intriguingly, abnormal colonies were noted to precede the rebound in bacterial growth

with reversion to normal morphology by 24 hours. Combination treatments exhibited a higher frequency of these abnormal colony variants compared to the bacteriophage alone treatments.

PHG001 impact on *E. coli* O157:H7 growth and *stx* gene expression

Several triggers, including antibiotics and other bacteriophages, are known to increase the production of Shiga-like toxin (*stx*) by *E. coli* O157:H7. To determine the effect of PHG001 on this important toxin, we examined the expression of *stx2c* in *E. coli* O157:H7 strain TW14359. Mitomycin C, a known inducer of *stx*, served as a positive control for increased *stx2c* expression. PHG001 ($1x10^8$ PFU/ml), mitomycin C (10 μg/ml), or a combination of both was added to exponentially growing *E. coli* O157:H7. Bacterial growth followed the trends observed with ampicillin. A 1-fold decrease in both the mitomycin C and PHG001 cultures was observed by hour 2, but a more substantial 3-fold decrease was observed at hour 3 for the cultures treated with both the PHG001 alone and the combination of PHG001 + mitomycin C (Figure. 4.9A). The mitomycin C treated cultures had a 1-fold decrease in *stx2c* expression compared to the untreated O157:H7 strain after 3 hours. Because three hours of post- bacteriophage infection correlated with the start of the exponential phase for PHG001-resistant bacteria, we assessed *stx2c* expression at this timepoint. Mitomycin C treated cultures had an 18-fold increase in *stx2c* expression, while the bacteriophage treated cultures exhibited a 16-fold reduction in expression relative to the untreated culture (Figure 4.9B). Moreover, the combination of bacteriophage and mitomycin C demonstrated a 4-fold reduction in *stx2c* expression.

DISCUSSION

Herein, we determined the composition of the isolated virus communities among patients with enteric infections (cases) and healthy individuals (controls) and characterized three bacteriophages. The most abundant viruses present in the virus communities were Caudovirales (Siphoviridae, Podoviridae), and Microviridae, which have been reported as dominant members of the virome (2, 3, 6, 42). Anelloviridae, a single-stranded virus that has been reported to be elevated in disease (43–48), was above the study average in 57% of the viromes isolated from cases. Although Anelloviridae is ubiquitous and has been classified as a commensal virus not directly linked to disease (49), an association has been found between nasal Anelloviridae loads and bronchial inflammation (50). Since only three healthy control samples were available to compare to the case samples, our ability to detect differences by case status was limited.

Nonetheless, cases had an increased abundance of Poxviridae, a viral family containing members that can cause human infections. Poxviridae abundance has previously been shown to be increased in patients with gestational diabetes (51) and HIV (52), though it was reported as a false-positive because a virus-only database was utilized, leading to an incorrect annotation. Despite using the entire non-redundant protein database in NCBI in this study, it is still possible that Poxviridae was incorrectly annotated, thereby leading to false estimates of increased abundance. Cross-assembly of the reads specific to Poxviridae, which is a computationally intensive process, is needed in future studies for confirmation. Indeed, we estimate that it will take 200 computer hours to compile the reads representing Poxviridae. This endeavor would be worthwhile as the results would provide insight into different fields. If Poxviridae genomes are definitively present within the viromes, then this study would represent the first reported incident of Poxviridae in cases of acute bacterial gastroenteritis. If the presence of Poxviridae is falsely annotated, however,

then this will provide information to bioinformaticians regarding optimization of the methodology used for annotation. If the genome that is being annotated as Poxviridae is novel, then further characterization of the viral genome needs to be performed as was done previously for cross-assembly phage (53), which has not been characterized phenotypically. Cross-assembly phage represents one of the most abundant proposed bacteriophage families found in human fecal samples (54). At the time of our analysis, the database (NCBI nr-protein) used for annotation did not contain cross-assembly phage, so these were not evaluated herein. Because many of the viral sequences were unannotated (83% on average across all samples), these sequences represent "viral dark matter" as was suggested previously (2, 3, 6, 55). More comprehensive databases are therefore needed to study the unknown viruses present. Additionally, this study focused solely on DNA viruses; future work should utilize reverse transcriptase to study the RNA component of the virome as well.

We also tested the functionality of isolated VLP communities from healthy and sick individuals and demonstrated that these communities could inhibit the growth of three types of pathogenic enteric bacteria at higher frequencies compared to three commensal *E. coli* strains (78% and 39%, respectively). While there is limited research that compares the effect of VLPs on commensal and pathogenic bacteria, similar results for individual bacteriophage s were observed *in vivo* using a mouse model. Kasmanet. *et al*, for instance, found that 22 commensal *E. coli* strains were resistant to 59% of single bacteriophage challenges with lambda, M13, P1, T4, T7, and PhiX174 coliphage (56). It is indeed possible that there are distinct genotypic or phenotypic differences between pathogenic and commensal bacteria that allow them to ward off bacteriophage infection. Commensal bacteria are exposed to resident gut bacteriophage populations frequently, whereas many enteric pathogens are transient. Therefore, pathogenic bacteria may not necessarily

have the resistance that commensal bacteria need to protect against repeated bacteriophage infections. This difference may explain why pathogenic bacteria were more likely to be infected and lysed compared to commensals, though a future study with more strains and viral communities should be conducted for confirmation. Studies have also shown that bacteriophage treatment may not affect the resident microbiota. For instance, *Bacteroides* and *Lactobacillus*, essential members of a healthy microbiome, were not affected by the use of a bacteriophage for the treatment of *C. difficile* (57), which is because this bacteriophage was restricted to specific hosts.

Three different bacteriophage s (PHG001, PHG002, PHG003) were isolated from the viral communities following infection of *E. coli* O157:H7. Two of these bacteriophage s, PHG002 and PHG003, represent related lysogenic bacteriophage s that were found in the 15 viromes examined, and in 23 additional *E. coli* genomes. Lysogenic bacteriophage s are important for pathogen evolution as they often carry genes such as virulence factors and antibiotic resistance genes, which can be transferred to other bacteria via horizontal gene transfer. We determined that both PHG002 and PHG003 harbor an exonuclease encoded by *sbcC*. SbcC along with SbcD, has been shown to process DNA intermediates at the convergence sites of replication forks, which allows for normal chromosome replication (58). Deletion mutations in *sbcC* and *sbcD* led to incomplete replication and genomic instability (58), while base-pair mutations in *sbcC* were shown to increase mitomycin C sensitivity (59). It is, therefore, possible that an additional copy of *sbcC* on a prophage integrated within the bacterial chromosome could enhance genomic stability in the host.

The lytic bacteriophage, PHG001, was found to be related to other bacteriophage s available in the NCBI database but was classified as unique based on the VipTree phylogeny. Importantly, PHG001 successfully inhibited the growth of three *E. coli* O157:H7 (TW14359, TW14313, TW18502) strains but allowed commensal strains (TW17000, TW17041) to grow

uninhibited. A prior study demonstrated that bacteriophage isolated from dairy and cattle feedlot manure could target and lyse STEC serogroups O26, O111, and O157 with high frequency (60). Indeed, PHG001 exhibited a similar host profile and could inhibit all seven O157:H7 strains examined as well as those belonging to O26 and O103. Commensal *E. coli* belonging to serogroups O6, O7, and O2 were also inhibited, suggesting variation in the inhibition potential of specific bacteriophage s across serogroups. These findings are supported by data from other studies showing that different O157:H7-lysing phages have wide host ranges (38, 61, 62). The significance of a broader host range means additional opportunities for phage infection and replication, which could allow the virus to become a resident within a given microbial environment.

The development of resistance to bacteriophage infection is a common occurrence and has been well studied. It was hypothesized that within a co-culture of bacteriophage and host bacteria, the bacteriophage would infect a subpopulation of bacteria, and ultimately the replication rate of the bacteriophage becomes tied to the subpopulation of host it can infect (1, 63). Resistance can arise due to mutations or possibly differences in transcription. For these reasons, multiple bacteriophages have been used to overcome bacterial resistance to a single bacteriophage. Indeed, previous work has shown that bacteriophage cocktails can drastically improve the efficacy of bacteriophage treatment. In one study, the combination of three bacteriophages isolated from human fecal samples in a cocktail demonstrated a five-fold reduction in *E. coli* O157:H7 concentration (64). Another study of three different bacteriophages, which were pooled into a bacteriophage cocktail, demonstrated a 5-log reduction in *E. coli* O157:H7 growth (65). Comparatively, we demonstrated that PHG001 reduced the concentration of *E. coli* O157:H7 by three-fold over 5 hours, however, a rebound in bacterial growth was observed by 24 hours. Such rebounds, as measured by the turbidity of the bacterial culture (65), have been described for *E. coli*

O157:H7 and suggest that bacteriophage resistance is common. Future studies should, therefore, focus on the inclusion of multiple bacteriophages targeting *E. coli* O157:H7 in combination with PHG001 to determine whether *E. coli* O157:H7 can overcome infection by multiple bacteriophage. The impact of bacteriophage on bacterial cells should also be evaluated given that abnormally shaped resistant colonies were observed in our study and others (65, 66). Indeed, it was suggested that these abnormal colonies have deficiencies in the structure of the cell wall, or O-antigen, as the colony morphology has a rough appearance (67). Such deficiencies could enhance the ability of some antibiotics to enter the bacterial cell, resulting in lysis and a synergistic effect, which was observed herein.

It is important to note that the abnormal *E. coli* O157:H7 colony morphology observed in our study was dependent on the presence of a bacteriophage. The highest frequency of irregular colonies appeared in the hour preceding and at the point of a bacteria rebound following PHG001 infection (data not shown). The abnormal colonies were similar to those recovered from mice and steer samples treated with O157:H7 specific bacteriophages (68). The abundance of abnormal colonies, however, was at much lower concentrations (<30 CFU/ml) when isolated directly from animals than in our experiments, which resulted in at least $1x10^6$ CFU/ml. We observed resistance, or a rebound in bacterial growth, within hours, whereas sample timing in the prior study found bacteriophage resistance to occur in O157:H7 at day seven (68). These findings suggest that there is variation in resistance development across strains and bacterial isolates. We found that resistance persists with subsequent culturing of resistant colonies, while a prior study found that resistance was lost with subsequent culture and growth of a resistant colony (68).

Abnormal colonies with different morphologies have also been observed following bacteriophage infection (69); these colonies are similar to small colony variants that form

following exposure to antibiotics and are linked to drug resistance (70). While these colonies are distinct, their growth kinetics in co-culture was similar to those observed in the experiments involving PHG001 infection of *E. coli* O157:H7 strain TW14359. A previous study demonstrated that a bacteriophage challenge resulted in a rapid decrease in *E. coli* O157:H7 concentration followed by a subsequent rebound and plateau of the bacteriophage population (71). Resistance was proposed to be due to an alteration in the outer-membrane or LPS structure (69). Further characterization of the physiology and abundance of these resistant colonies needs to be evaluated to understand their role in bacteriophage resistance and pathogenesis.

This study also examined Shiga toxin (*stx*) gene expression, a marker for STEC virulence and infection in humans as Stx production is a crucial contributor to hemorrhagic colitis, bloody diarrhea, and hemolytic uremic syndrome (72, 73). STEC infections result in a mortality rate between 3% to 5% (72) with 20% of surviving patients developing permanent kidney dysfunction (73). STEC harbors *stx* genes located on at least one bacteriophage incorporated into the STEC genome as a prophage, which, when the bacteriophage is replicated, causes cell lysis and Stx production (74). The genes that encode Stx are carried on lambdoid bacteriophage s and can be easily transferred to pathogenic and commensal strains of Enterobacteriaceae (75–77); *stx* expression is controlled by the bacteriophage repressor (78–80). Stx prophages can undergo spontaneous phage induction and enter the lytic cycle (81). Activation of the SOS response through DNA damage or halting DNA replication leads to replication of Stx bacteriophages (82). This bacteriophage induction occurs through Rec-A mediated cleavage of the bacteriophage repressor leading to toxin expression and bacteriophage replication (74, 82, 83). Numerous stressors can induce the SOS response, including UV light (84), hydrogen peroxide (85), EDTA (86), antibiotics (87) such as mitomycin C (83), and bacteriophage infection (74, 88). Bacteriophage infection can

cause changes in bacterial cellular processes, which includes altering the normal replication of the bacterium (89).

Bacteriophage can increase the amount of single-stranded DNA (90) or directly degrade the bacterial chromosome (91, 92), inducing a stress response. Studies have shown that Stx prophages are more prone to induction relative to other prophages (81). Given these results from the literature, we hypothesized that PHG001 would increase *stx* expression by inducing prophage-mediated *recA* transcription by interfering with host DNA replication. Surprisingly, PHG001 reduced *stx* expression alone and in combination with mitomycin C, which has been shown to induce toxin expression (25). Indeed, *stx* expression was increased in *E. coli* O157:H7 by 18-fold in the presence of mitomycin C relative to basal expression. PHG001 can negate the induction of the SOS response in STEC as toxin expression was reduced 16-fold in culture alone following normalization to a constitutively expressed housekeeping gene, *gyrA*. Bacterial gene expression decreases globally with a lytic bacteriophage infection (93), though we observed a differential decrease in *stx* expression that would not have been observed if it was due solely to the global decrease by bacteriophage infection.

Lytic bacteriophage has also been shown to increase the amount of resident prophage DNA present in bacterial cells after infection (89), which, if this occurred in this study, would result in an increase in toxin expression due to co-expression of *stx* and the stx prophage. A prior study performed transcriptomics to study the regulatory role of integrated bacteriophage in *stx*-positive *E. coli* (94) and lytic bacteriophage (93). After infection, lytic bacteriophage was shown to take over host transcriptional machinery in order to produce more bacteriophage (93). In *Pseudomonas aeruginosa*, for example, a lytic bacteriophage was shown to suppress the transcription of a resident prophage, P2 (93). These findings support the notion that prophage expression can be

affected by a lytic bacteriophage. In this study, repressing Stx prophage expression would subsequently result in a decrease *stx* expression because they are co-expressed. Nonetheless, transcriptomic studies are needed to expand on this work and determine the mechanism of *stx* expression inhibition by PHG001. Similar effects of reduced toxin production in the presence of an exogenous bacteriophage have been noted with *Clostridium difficile* in batch fermentation (57).

Importantly, the combination of PHG001 and ampicillin completely inhibited the growth of *E. coli* O157, thereby representing a synergistic effect between bacteriophage and antibiotics. Phage and antibiotics have been shown to have this effect with a wide range of bacteriophage and antibiotics against multiple bacterial species. For instance, bacteriophage plus β-lactam antibiotics and quinolones were effective against uropathogenic *E. coli* (95), while bacteriophage and gentamicin could impact *Staphylococcus aureus* (96) and bacteriophage and tobramycin were useful against *E. coli* and *Pseudomonas aeruginosa* (97). Different antibiotics may have different effects on *stx* production. Ampicillin, for example, has been shown to have minimal impact on Stx production (87); this finding is likely because ampicillin acts in an SOS-independent manner, which if combined with bacteriophage, could result in an even more drastic reduction in toxin production. Bacteriophage holins and antibiotics work in conjunction to permit more extensive bacteriophage replication and subsequent host lysis (98). Additional studies related to this work, however, should quantify the bacteriophage concentration in conjunction with the antibiotics tested to observe changes in bacteriophage concentration as it relates to bacteria growth. Further research will need to consider the possibilities of phage-antibiotic synergy. The inhibition of toxin expression could provide an avenue for further investigation that could be beneficial for human health. It is important to note that PHG001 was found in 87% of the metagenomes present in the initial study and similar bacteriophage have been found in animal models (68).

APPENDIX

Table 4.1. The effect of intestinal viral-like particles (VLPs) on lysis of three bacterial pathogens and three commensal *Escherichia coli* strains

			Pathogens			Commensal *E. coli*		
ID	Stool source	Type of infection	*Shigella sonnei* (TW16390)	*Salmonella* Typhimurium (TW16372)	*E. coli* O157:H7 (TW14359)	TW17000	TW17041	TW17368
ER644	Case	*C. jejuni*	+	+	+	-	-	-
ER629	Case	*C. jejuni*	+	+	+	-	-	-
ER641	Case	*C. jejuni*	-	+	+	-	-	+
ER649	Case	*C. jejuni*	+	-	+	+	+	+
ER631	Case	*S. enterica* spp.	+	+	+	-	+	+
ER676	Case	*S. enterica* spp.	+	+	-	+	+	+
ER628	Case	*S. enterica* spp.	+	+	-	-	-	-
ER646	Case	*S. enterica* spp.	+	+	+	-	+	-
ER694	Case	*Shigella* spp.	+	+	+	-	-	-
ER640	Case	*Shigella* spp.	+	+	+	-	+	-
ER653	Case	*Shigella* spp.	+	+	+	-	-	-
ER661	Case	*Shigella* spp.	-	+	-	+	+	-
ER680	Case	*E. coli* (STEC)	+	-	-	-	-	-
ER657	Case	*E. coli* (STEC)	-	-	+	-	-	-
		Total lysis by case VLPs	11 (78.6%)	11 (78.6%)	10 (71.4%)	3 (21.4%)	6 (42.9%)	4 (28.6%)
ER664	Control	No infection	+	-	+	-	-	+
ER693	Control	No infection	+	+	+	-	-	+
ER708	Control	No infection	+	+	+	+	+	+
ER689	Control	No infection	+	+	+	+	+	+
		Total lysis by control VLPs	4 (100%)	3 (75%)	4 (100%)	2 (50%)	2 (50%)	4 (100%)

STEC = Shiga toxin-producing *E. coli*

Table 4.2. Virome sequencing results and coverage

Study ID	Reads Paired-forward Total Count (Mbp)	Reads remaining after low-quality read removal Count (%)	Reads remaining after human-read removal Count (%)	Total Viral Reads annotated Count (%)	Nonpareil Coverage (%)
ER628	185565 (92.8)	153692 (82.8)	136702 (88.9)	3572 (2.6)	77.3
ER631	2238040 (1119)	78287 (3.5)	27532 (35.2)	1846 (6.7)	13.4
ER640	934953 (467.5)	410744 (43.9)	407064 (91.9)	20799 (5.1)	94.8
ER641	2457652 (1228.8)	1071369 (43.6)	29694 (2.8)	325 (1.1)	48.4
ER644	116776 (58.4)	43107 (36.9)	40366 (93.6)	5659 (14.0)	83.0
ER646	216413 (108.2)	95956 (44.3)	95874 (99.9)	3065 (3.2)	93.0
ER649	1027408 (513.7)	455652 (44.3)	452157 (92.8)	76856 (17)	96.2
ER653	149214 (74.6)	15283 (10.2)	7031 (46.0)	491 (7.0)	38.6
ER661	411435 (205.7)	132230 (32.1)	85089 (64.3)	4112 (4.8)	62.4
ER676	2900746 (1450.4)	498257 (17.2)	36432 (7.3)	1281 (3.5)	33.5
ER680	273780 (136.9)	186159 (68.0)	184823 (99.3)	119180 (64.5)	96.3
ER689*	369038 (184.5)	175885 (47.7)	173769 (98.8)	27383 (15.8)	89.9
ER693*	714100 (357.1)	301210 (42.2)	295244 (98.0)	28674 (9.7)	88.7
ER694	478812 (239.4)	172989 (36.1)	169547 (98.0)	148249 (87.4)	93.5
ER708*	1291317 (645.7)	564402 (43.7)	563659 (99.9)	11401 (2.0)	97.1

* Samples from healthy individuals (controls)

Table 4.3. Sequencing results and coverage estimates for three bacteriophages capable of inhibiting the growth of *Escherichia coli* O157:H7

Phage ID	Type	Paired-forward total reads Count (Mbp)	After low-quality read removal Count (%)	Assembly Length (bp)	Reads Mapped Count (%)	Assembly Depth
PHG001	Lytic	500184 (250.1)	429312 (85.8%)	114632	428634 (99.8%)	934.8x
PHG002	Lysogenic	428774 (214.4)	349282 (81.4%)	32067	290472 (83.2%)	2264.6x
PHG003	Lysogenic	367658 (183.8)	286807 (78.0%)	32178	253590 (69.0%)	1977.0x

Table 4.4. Characteristics of the strains used to determine the host range of a novel lytic bacteriophage, PHG001

Accession Number	Strain	Strain type	Spot test result*	O-type	H-antigen
TW02054	ECOR-35	commensal	-	1	NM
TW02051	ECOR-26	commensal	-	104	21
TW03288	ECOR-28	commensal	-	104	2
TW02049	ECOR-24	commensal	-	15	NM
TW03299	ECOR-49	commensal	-	2	NM
TW03313	ECOR-61	commensal	+	2	NM
TW02062	ECOR-51	commensal	-	25	-
TW03279	ECOR-15	commensal	-	25	NM
TW03308	ECOR-54	commensal	-	25	1
TW02064	ECOR-59	commensal	-	4	40
TW03307	ECOR-53	commensal	+	4	-
TW03276	ECOR-10	commensal	+	6	10
TW03310	ECOR-56	commensal	+	6	1
TW02046	ECOR-12	commensal	+	7	32
TW02056	ECOR-38	commensal	+	7	NM
TW02057	ECOR-41	commensal	-	7	NM
TW03294	ECOR-39	commensal	+	7	NM
TW02055	ECOR-36	commensal	-	79	25
TW03272	ECOR-05	commensal	-	79	NM
TW03274	ECOR-08	commensal	-	86	NM
TW02058	ECOR-42	commensal	-	N	26
TW02059	ECOR-43	commensal	-	N	-
TW03273	ECOR-06	commensal	-	N	NM
TW03275	ECOR-09	commensal	-	N	NM
TW03278	ECOR-13	commensal	+	N	-
TW03315	ECOR-63	commensal	+	N	NM
TW03268	ECOR-01	commensal	-	NT	NM
TW03269	ECOR-02	commensal	-	NT	32

Table 4.4. (cont'd)

TW03271	ECOR-04	commensal	+	NT	NM
TW18515	STEC	pathogen	+	O103	H2
TW18525	STEC	pathogen	+	O103	H2
TW18531	STEC	pathogen	-	O103	H2
TW18538	STEC	pathogen	-	O103	H2
TW19067	STEC	pathogen	-	O103	H2
TW19078	STEC	pathogen	-	O103	H2
TW19079	STEC	pathogen	-	O103	H2
TW19085	STEC	pathogen	-	O103	H2
TW18511	STEC	pathogen	-	O111	H8
TW18523	STEC	pathogen	-	O111	H8
TW18524	STEC	pathogen	-	O111	H8
TW18527	STEC	pathogen	-	O111	H8
TW18990	STEC	pathogen	-	O111	H8
TW19035	STEC	pathogen	-	O111	H8
TW19051	STEC	pathogen	-	O111	H8
TW19057	STEC	pathogen	-	O111	H8
TW18526	STEC	pathogen	+	O26	H11
TW18535	STEC	pathogen	-	O26	H11
TW18536	STEC	pathogen	-	O26	H11
TW18585	STEC	pathogen	-	O26	H11
TW19056	STEC	pathogen	+	O26	H11
TW19068	STEC	pathogen	+	O26	H11
TW19070	STEC	pathogen	+	O26	H11
TW19088	STEC	pathogen	-	O26	H11
TW18494	STEC	pathogen	-	O45	H2
TW18496	STEC	pathogen	-	O45	H2
TW18504	STEC	pathogen	-	O45	H2
TW18505	STEC	pathogen	-	O45	H2

Table 4.4. (cont'd)

TW19074	STEC	pathogen	-	O45	H2
TW19076	STEC	pathogen	-	O45	H2
TW19080	STEC	pathogen	-	O45	H2
TW19083	STEC	pathogen	-	O45	H2
TW14588	STEC	pathogen	+	O157	H7
TW14313	STEC	pathogen	+	O157	H7
TW11039	STEC	pathogen	+	O157	H7
TW18482	STEC	pathogen	+	O157	H7
TW18484	STEC	pathogen	+	O157	H7
TW18485	STEC	pathogen	+	O157	H7
TW14359	STEC	pathogen	+	O157	H7
TW16372	*Shigella sonnei*	pathogen	+		
TW16390	*Salmonella* Typhimurium	pathogen	-		
TW17000	*E. coli*	commensal	-		
TW17041	*E. coli*	commensal	-		
TW17368	*E. coli*	commensal	-		

ECOR = *E. coli* reference collection (35); STEC = Shiga toxin-producing *E. coli*

* + indicates growth inhibition, while – indicates no inhibition.

Note: The O- and H-antigen types are not known for commensal *E. coli* TW17000, TW17041, and TW171368, which were isolated from the stools of healthy participants.

Figure 4.1. Lysis of commensal and pathogens by intestinal virus-like particles (VLPs) Bacterial strains challenged with viral communities, or VLPs, isolated from 18 individuals. **A)** Lysis frequency; and **B)** average phage titer (Log plaque-forming units (PFU)/ml) per infection. Strains include three *Shigella sonnei* (TW16372), *Salmonella Typhimurium* (TW16390), STEC (TW14359), and three commensal *E. coli* (TW17000,17041,17368).

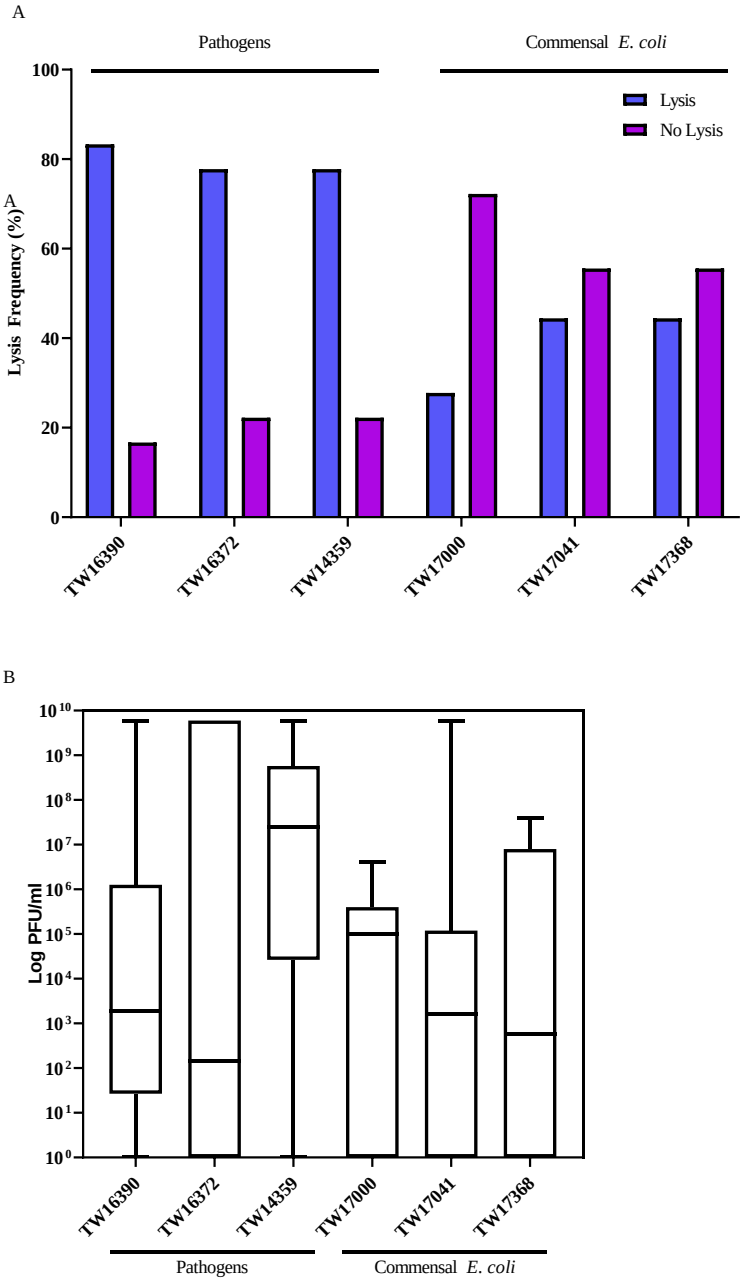
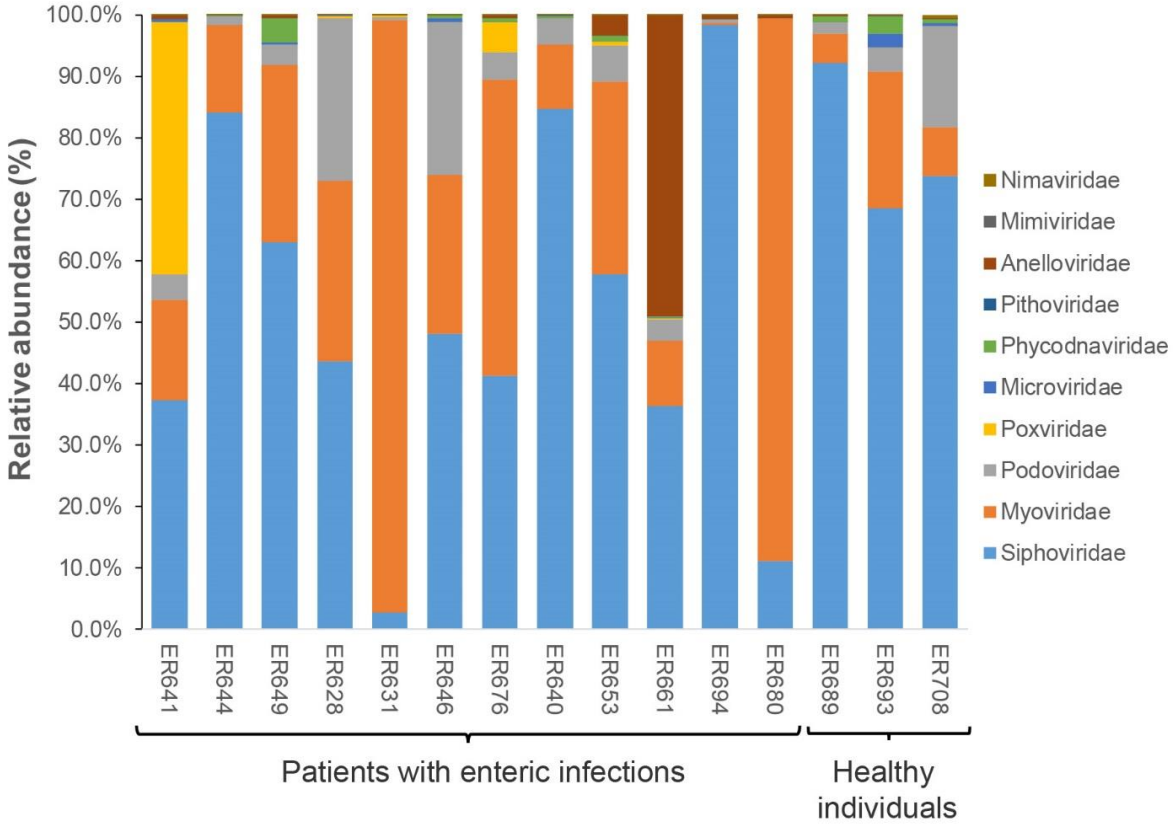


Figure 4.2. Viral community profiles isolated from the stools of patients with enteric infections and otherwise healthy participants The relative abundance of each viral family is shown across samples; count data was log-transformed, and total sum scaled.



Note: Only three stool communities from healthy individuals were available for analysis, thereby limiting our ability to examine differences by source.

Figure 4.3. Sequence analysis of lysogenic phages, PHG002 and PHG003, recovered following infection of Escherichia coli O157:H7 strain TW14359 A) A proteomic dendrogram constructed using VipTree shows the relationship between PHG002 and PHG003 (blue stars) and 23 closely related phage genomes. The virus family, Myoviridae (light green), and predicted host group, Gammaproteobacteria (dark green), are also indicated. Branch length scaling is represented linearly. B) Pairwise-alignment of five closely related phages identified in the proteomic tree. The dot plot (left) visualizes the comparison between the two viral genomes, while the blue bar represents the genome map of each virus. The percent identity, shown as the bar beneath each viral genome, indicates the similarity of the pair-wise comparison between the viruses.

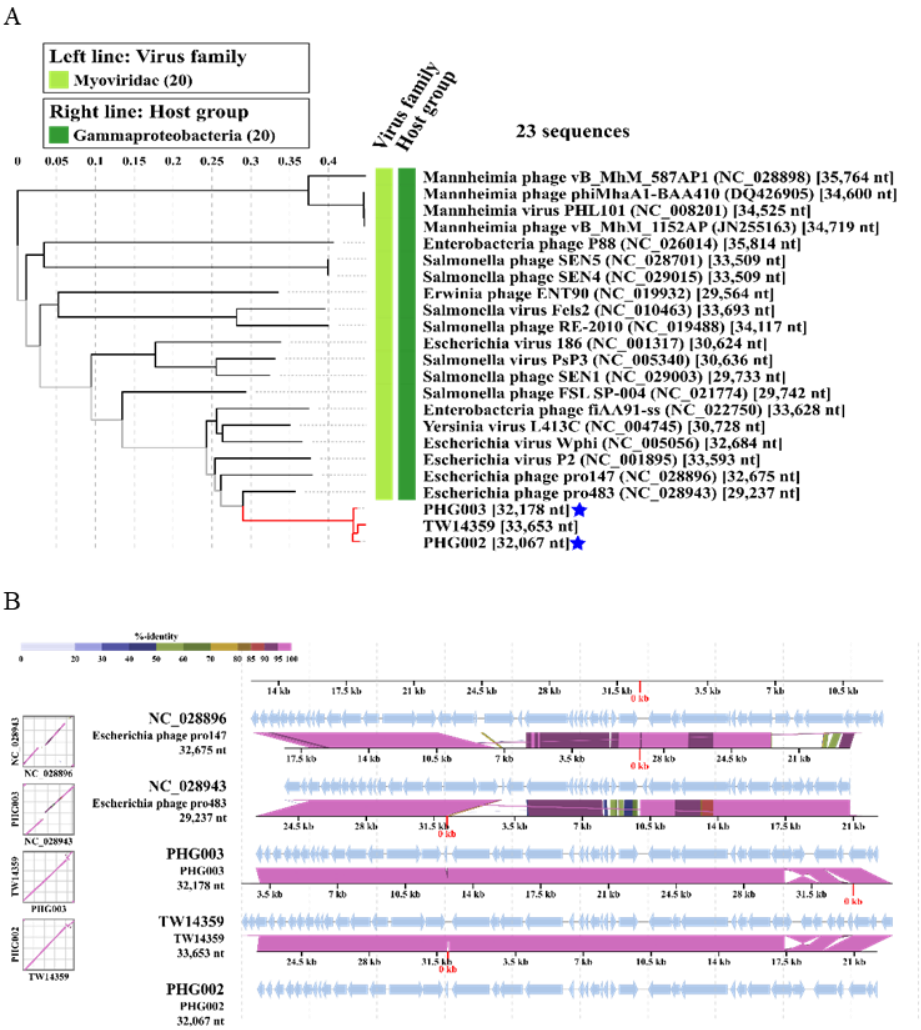


Figure 4.4. Neighbor-joining tree of BLAST alignments of PHG002 A neighbor-joining tree was constructed from BLAST alignments to the *Escherichia coli* genome database on NCBI. Listed genomes were found to possess PHG002 or a closely related variant of PHG002 in the genome (100% alignment and >99% percent identity). The vertical bar designates a cluster of the most closely related bacteriophages (indicated as a green triangle at the node). The samples outside of the cluster have a percent identity (<80%).

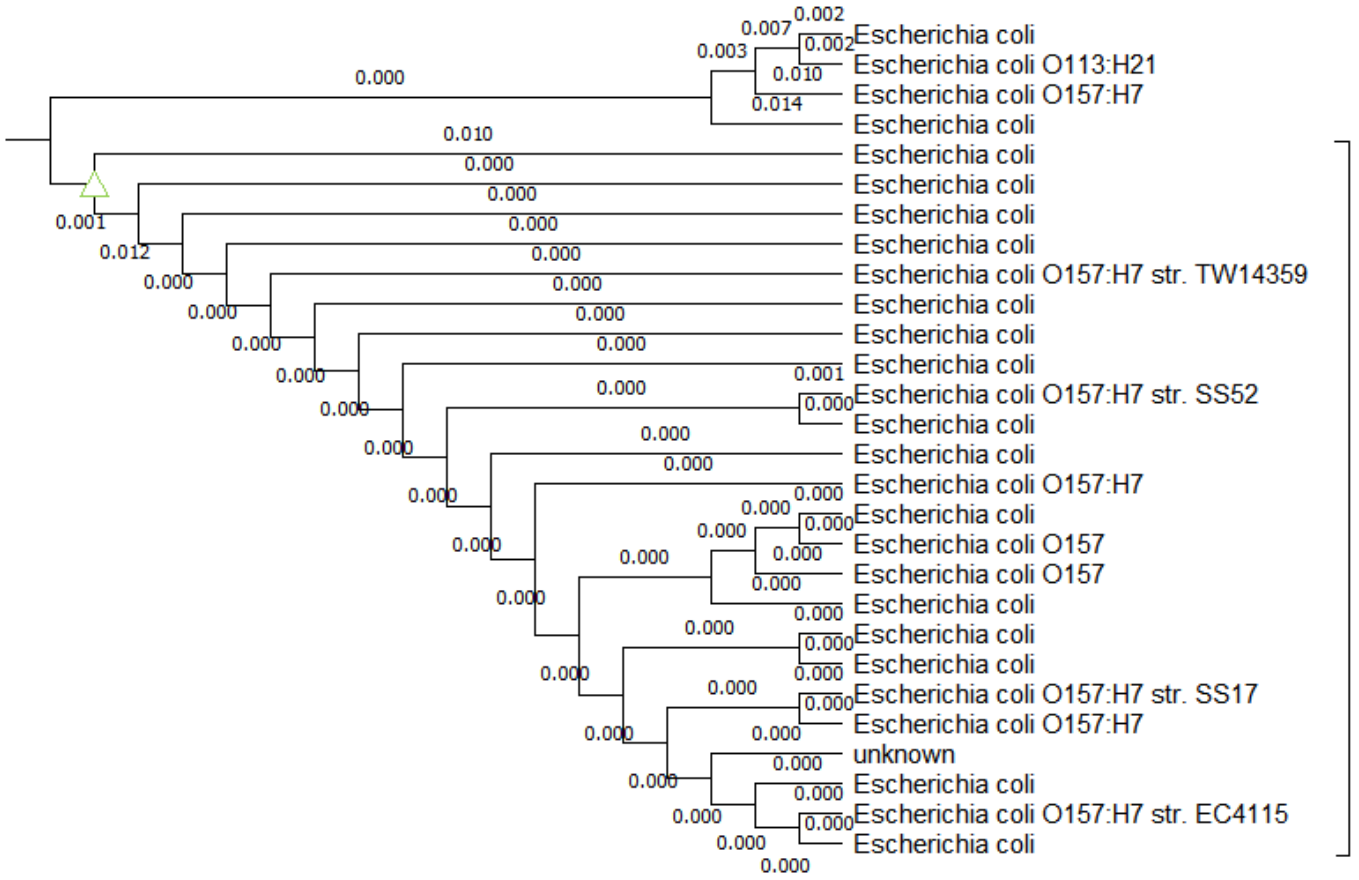


Figure 4.5. PHG001 genomic map The function of a given gene is represented in the legend at the bottom. The base pairs along the genome are represented by the lines, empty boxes represent predicted proteins of unknown function.

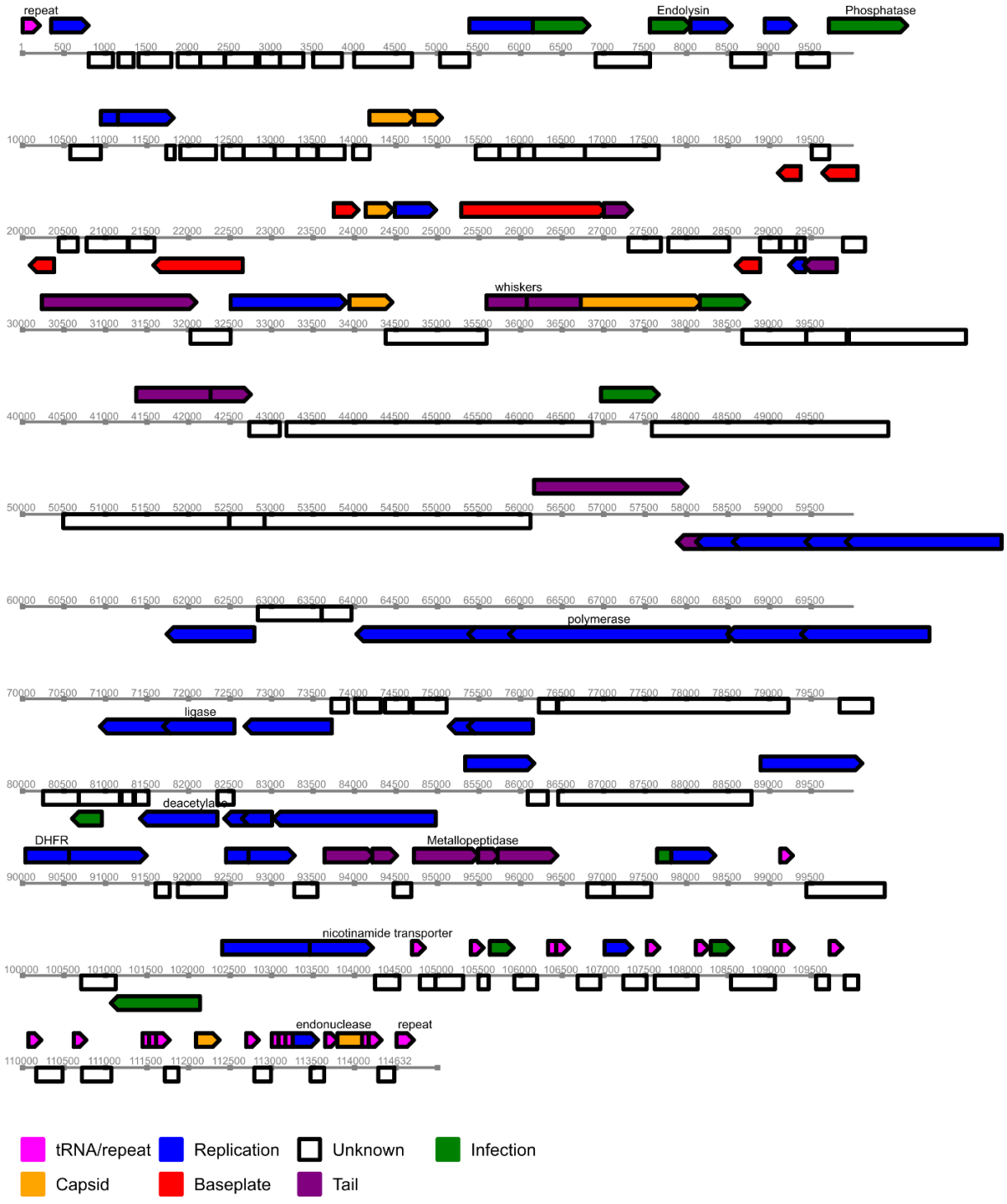


Figure 4.6. Sequence analysis of lytic phage PHG001 recovered following infection of Escherichia coli O157:H7 strain TW14359 A) A proteomic dendrogram constructed using VipTree shows the relationship between PHG001 (green star) and 23 closely related phage genomes. The virus family, Siphoviridae (orange), and predicted host group, Gammaproteobacteria (dark green) are indicated for each phage. Branch length scaling is linear. B) Pairwise-alignment of closely related phages identified in the proteomic tree. The dot plot (left) visualizes the comparison between two viral genomes. The blue bar represents the genome map of each virus. The percent identity, shown as the bar beneath each viral genome, indicates the pairwise similarity between the two-related viruses.

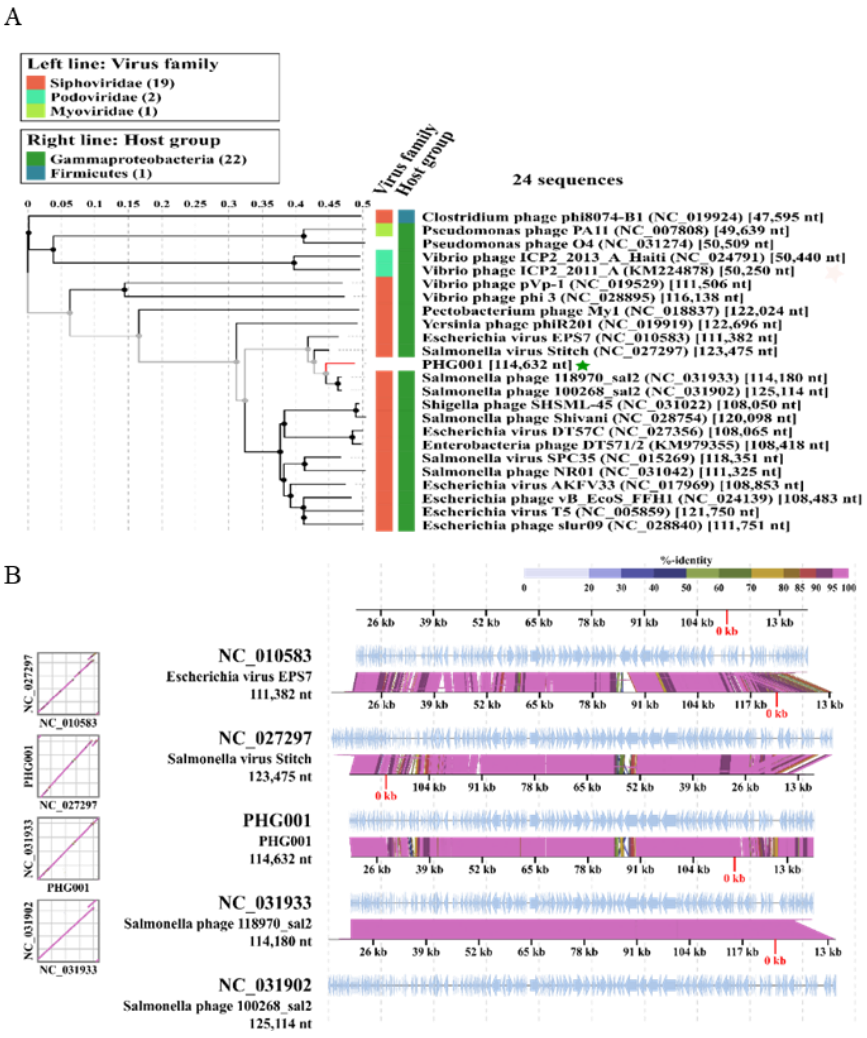


Figure 4.7. PHG001 growth in the Escherichia O157:H7 host The growth of PHG001 in plaque-forming units (PFU)/ml were evaluated in *E. coli* O157:H7 strain TW14359 over 24 hours. Timepoints were taken every 20 minutes for the first 2 hours, every hour up until hour 5, and then at hour 24 hours. Arrows represent latency periods. The initial burst size was calculated based on the time-points before and after the first burst. Experiments performed at a multiplicity of infection of 1. Error bars represent two standard deviations (N=3).

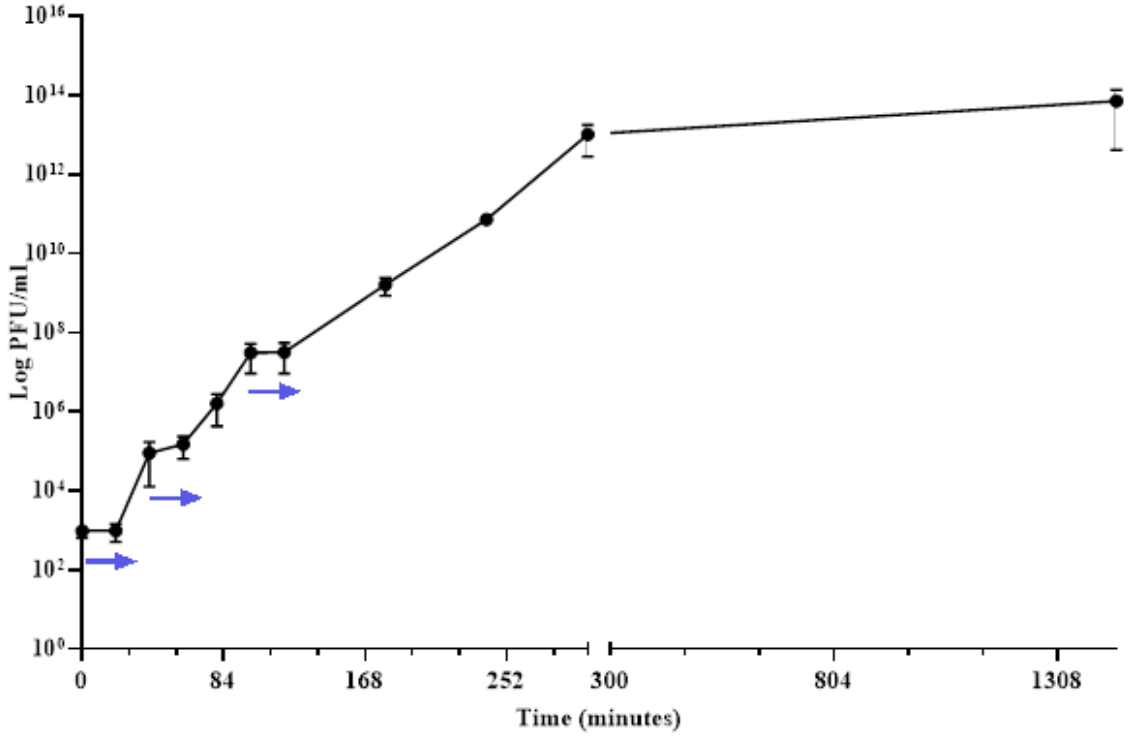


Figure 4.8. Effect of bacteriophage and ampicillin on the growth of Escherichia coli O157:H7

E. coli O157:H7 strain TW14313 was challenged with media (mock, orange line), bacteriophage (10^8 PFU/ml, red line), ampicillin (3.8 µg/ml, blue line) and a combination of bacteriophage (10^8 PFU/ml) and 3.8 µg/ml of ampicillin (purple line). The bacterial concentration was measured in CFU/ml every hour for up to 5 hours. Timepoint "PT" represents the concentration of the culture before inoculation, or pre-treatment with one of the three treatments, which were added to the bacterial culture at time 0. Error bars represent two standard deviations (N=3).

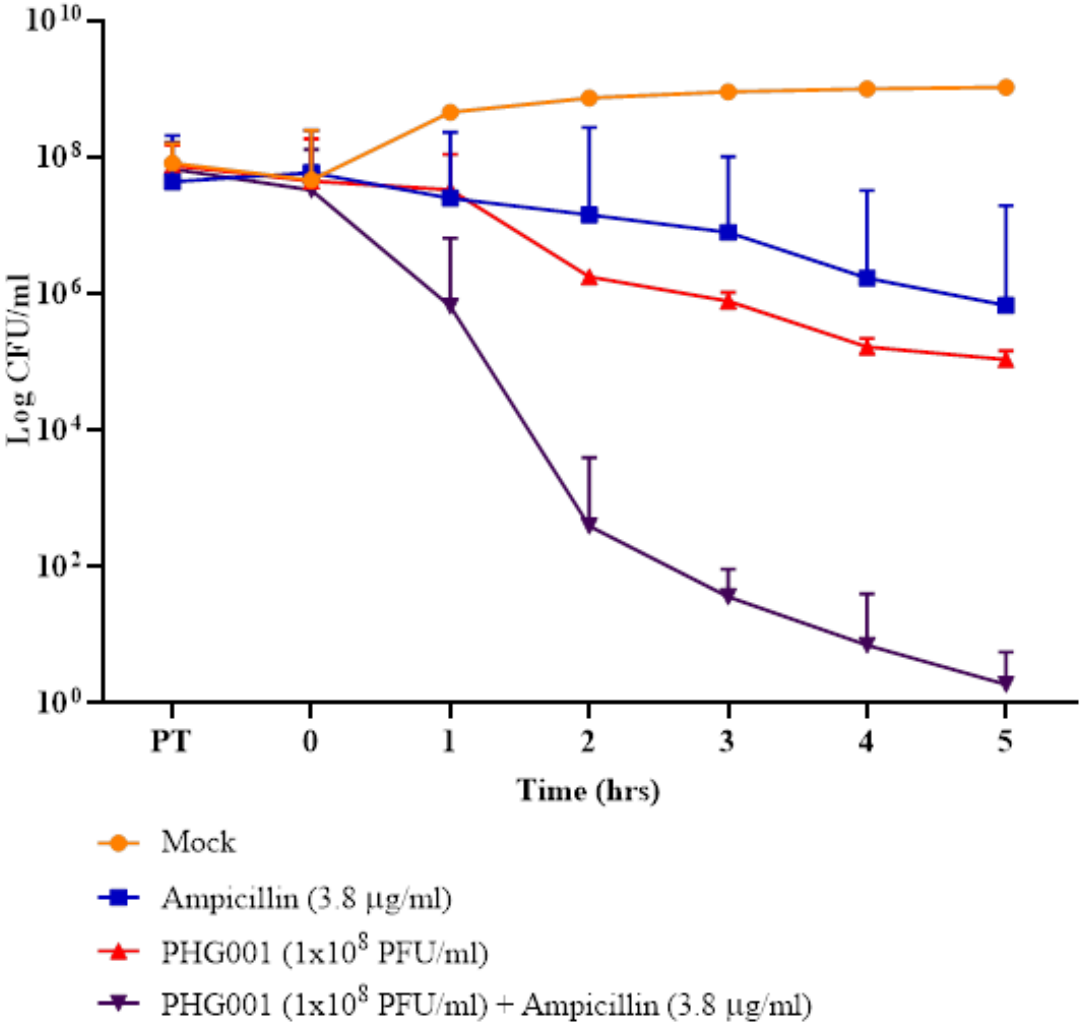
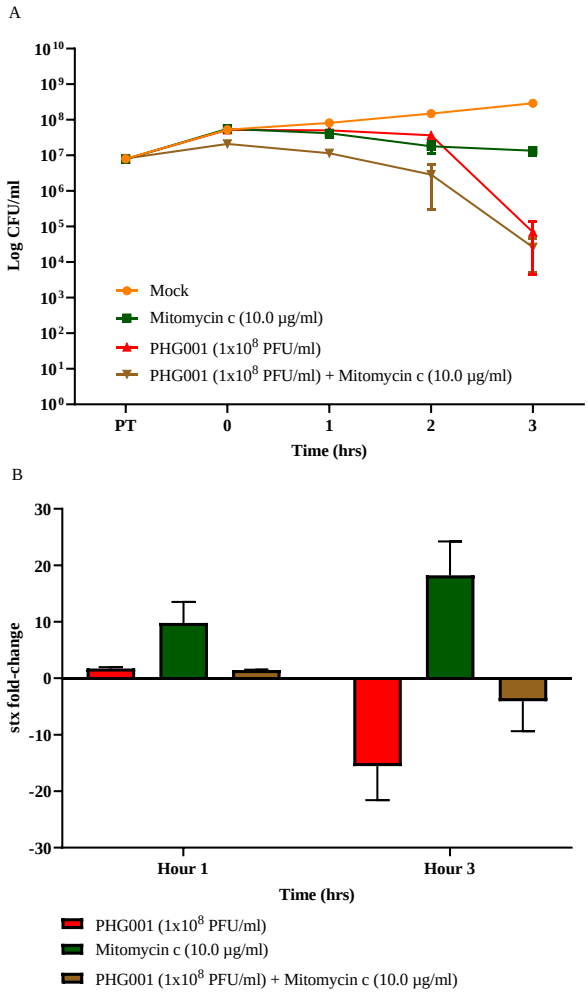


Figure 4.9. Effect of bacteriophage and mitomycin C on Escherichia coli O157:H7 growth and expression of Shiga toxin 2c A) *E. coli* O157:H7 strain TW14359 challenged with bacteriophage (10^8 PFU/ml, red line), mitomycin C (10.0 µg/ml, green line), and a combination of bacteriophage (10^8 PFU/ml) and mitomycin C (10.0 µg/ml, brown line) at time = 0. PT (pre-treatment) is the culture concentration 15 mins before the challenge. Error bars are two standard deviations. B) The fold-change in *stx2c* expression was measured at hour 3 and normalized to the level of expression observed in the untreated O157:H7 strain (Mock) using $2^{-\Delta\Delta CT}$ (40). Each box represents the 1st and 3rd quintiles between three biological replicates, while the whiskers represent the minimum and maximum, and the line is the median.



REFERENCES

REFERENCES

1. **Bohannan BJM**, **Lenski RE**. 2000. Linking genetic change to community evolution: Insights from studies of bacteria and bacteriophage. Ecol Lett.

2. **Breitbart M**, **Hewson I**, **Felts B**, **Mahaffy JM**, **Nulton J**, **Salamon P**, **Rohwer F**. 2003. Metagenomic analyses of an uncultured viral community from human feces. J Bacteriol **185**:6220–6223.

3. **Reyes A**, **Haynes M**, **Hanson N**, **Angly FE**, **Heath AC**, **Rohwer F**, **Gordon JI**. 2010. Viruses in the faecal microbiota of monozygotic twins and their mothers. Nature **466**:334–338.

4. **Kidambi SP**, **Ripp S**, **Miller R V.** 1994. Evidence for phage-mediated gene transfer among Pseudomonas aeruginosa strains on the phylloplane. Appl Environ Microbiol **60**:496–500.

5. **Oliver KM**, **Degnan PH**, **Hunter MS**, **Moran NA**. 2009. Bacteriophages encode factors required for protection in a symbiotic mutualism. Science (80-) **325**:992–994.

6. **Minot S**, **Sinha R**, **Chen J**, **Li H**, **Keilbaugh SA**, **Wu GD**, **Lewis JD**, **Bushman FD**. 2011. The human gut virome: Inter-individual variation and dynamic response to diet. Genome Res **21**:1616–1625.

7. **Dicksved J**, **Halfvarson J**, **Rosenquist M**, **Järnerot G**, **Tysk C**, **Apajalahti J**, **Engstrand L**, **Jansson JK**. 2008. Molecular analysis of the gut microbiota of identical twins with Crohn's disease. ISME J **2**:716–727.

8. **Turnbaugh PJ**, **Hamady M**, **Yatsunenko T**, **Cantarel BL**, **Duncan A**, **Ley RE**, **Sogin ML**, **Jones WJ**, **Roe BA**, **Affourtit JP**, **Egholm M**, **Henrissat B**, **Heath AC**, **Knight R**, **Gordon JI**. 2009. A core gut microbiome in obese and lean twins. Nature.

9. **Stern A**, **Mick E**, **Tirosh I**, **Sagy O**, **Sorek R**. 2012. CRISPR targeting reveals a reservoir of common phages associated with the human gut microbiome. Genome Res **22**:1985–1994.

10. **Barr JJ**, **Auro R**, **Furlan M**, **Whiteson KL**, **Erb ML**, **Pogliano J**, **Stotland A**, **Wolkowicz R**, **Cutting AS**, **Doran KS**, **Salamon P**, **Youle M**, **Rohwer F**. 2013. Bacteriophage adhering to mucus provide a non-host-derived immunity. Proc Natl Acad Sci U S A **110**:10771–6.

11. **Singh P**, **Teal TK**, **Marsh TL**, **Tiedje JM**, **Mosci R**, **Jernigan K**, **Zell A**, **Newton DW**, **Salimnia H**, **Lephart P**, **Sundin D**, **Khalife W**, **Britton RA**, **Rudrik JT**, **Manning SD**. 2015. Intestinal microbial communities associated with acute enteric infections and disease recovery. Microbiome **3**:45.

12. **Yamamoto R**, **Alberts M**. 1970. Rapid Bacteriophage Glycol Sedimentation Its

Application in the to Virus of Polyethylene Purification. Virology **744**:734–744.

13. **Manning SD**, **Motiwala AS**, **Springman AC**, **Qi W**, **Lacher DW**, **Ouellette LM**, **Mladonicky JM**, **Somsel P**, **Rudrik JT**, **Dietrich SE**, **Zhang W**, **Swaminathan B**, **Alland D**, **Whittam TS**. 2008. Variation in virulence among clades of Escherichia coli O157:H7 associated with disease outbreaks. Proc Natl Acad Sci.

14. **Singh P**, **Sha Q**, **Lacher DW**, **Del Valle J**, **Mosci RE**, **Moore JA**, **Scribner KT**, **Manning SD**. 2015. Characterization of enteropathogenic and Shiga toxin-producing Escherichia coli in cattle and deer in a shared agroecosystem. Front Cell Infect Microbiol **5**:29.

15. **Centers for Disease Control and Prevention**. 2015. Introduction to Epi Info ™ 7 Using Epi Info ™. Epi Info.

16. **Cormier J**, **Janes M**. 2014. A double layer plaque assay using spread plate technique for enumeration of bacteriophage MS2. J Virol Methods.

17. **Rodriguez-R LM**, **Gunturu S**, **Tiedje JM**, **Cole JR**, **Konstantinidis KT**. 2018. Nonpareil 3: Fast Estimation of Metagenomic Coverage and Sequence Diversity. mSystems.

18. **Bolger a. M**, **Lohse M**, **Usadel B**. 2014. Trimmomatic: A flexible read trimming tool for Illumina NGS data. Bioinformatics **30**:2114–2120.

19. **Langmead B**, **Salzberg SL**, **Langmead**. 2013. Bowtie2. Nat Methods **9**:357–359.

20. **Menzel P**, **Ng KL**, **Krogh A**. 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun **7**:11257.

21. **Palarea-Albaladejo J**, **Martín-Fernández JA**. 2015. ZCompositions - R package for multivariate imputation of left-censored data under a compositional approach. Chemom Intell Lab Syst.

22. **The R Foundation**. 2019. The R Project for Statistical Computing. 3.6.1. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, Vienna, Austria.

23. **Wickham H**. 2006. An introduction to ggplot: An implementation of the grammar of graphics in R. Statistics (Ber).

24. **Oksanen J**, **Kindt R**, **Legendre P**, **O'Hara B**, **Simpson GL**, **Solymos PM**, **Stevens MHH**, **Wagner H**. 2007. The vegan package. Community Ecol Packag.

25. **Andrews S**. 2010. FastQC: A quality control tool for high throughput sequence data. babraham Bioinforma http://www.bioinformatics.babraham.ac.uk/projects/.

26. **Bankevich A**, **Nurk S**, **Antipov D**, **Gurevich AA**, **Dvorkin M**, **Kulikov AS**, **Lesin VM**, **Nikolenko SI**, **Pham S**, **Prjibelski AD**, **Pyshkin A V.**, **Sirotkin A V.**, **Vyahhi N**, **Tesler G**, **Alekseyev M a.**, **Pevzner P a.** 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol **19**:455–477.

27. **Seemann T**. 2014. Prokka: Rapid prokaryotic genome annotation. Bioinformatics **30**:2068–2069.

28. **Aziz RK**, **Bartels D**, **Best AA**, **DeJongh M**, **Disz T**, **Edwards RA**, **Formsma K**, **Gerdes S**, **Glass EM**, **Kubal M**, **Meyer F**, **Olsen GJ**, **Olson R**, **Osterman AL**, **Overbeek RA**, **McNeil LK**, **Paarmann D**, **Paczian T**, **Parrello B**, **Pusch GD**, **Reich C**, **Stevens R**, **Vassieva O**, **Vonstein V**, **Wilke A**, **Zagnitko O**. 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics **9**:75.

29. **Altschul SF**, **Gish W**, **Miller W**, **Myers EW**, **Lipman DJ**. 1990. Basic local alignment search tool. J Mol Biol **215**:403–10.

30. **Pruitt K**, **Brown G**, **Tatusova T**, **Maglott D**. 2002. The Reference Sequence (RefSeq) Database. NCBI Handb 1–24.

31. **Arndt D**, **Marcu A**, **Liang Y**, **Wishart DS**. 2017. PHAST, PHASTER and PHASTEST: Tools for finding prophage in bacterial genomes. Brief Bioinform.

32. **Darling ACE**, **Mau B**, **Blattner FR**, **Perna NT**. 2004. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. Genome Res **14**:1394–1403.

33. **Rohwer F**, **Edwards R**. 2002. The phage proteomic tree: A genome-based taxonomy for phage. J Bacteriol.

34. **Nurk S**, **Meleshko D**, **Korobeynikov A**, **Pevzner PA**. 2017. metaSPAdes: a new versatile metagenomic assembler (Supplementary Material). Genome Res **27**:824–834.

35. **Ochman H**, **Selander RK**. 1984. Standard reference strains of Escherichia coli from natural populations. J Bacteriol.

36. **Gould LH**, **Mody RK**, **Ong KL**, **Clogher P**, **Cronquist AB**, **Garman KN**, **Lathrop S**, **Medus C**, **Spina NL**, **Webb TH**, **White PL**, **Wymore K**, **Gierke RE**, **Mahon BE**, **Griffin, for the Emerging Infection PM**. 2013. Increased Recognition of Non-O157 Shiga Toxin–Producing Escherichia coli Infections in the United States During 2000–2010: Epidemiologic Features and Comparison with E. coli O157 Infections. Foodborne Pathog Dis **10**:453–460.

37. **Mukherjee S**, **Mosci RE**, **Anderson CM**, **Snyder BA**, **Collins J**, **Rudrik JT**, **Manning SD**. 2017. Antimicrobial drug–resistant Shiga toxin–producing Escherichia coli infections, Michigan, USA. Emerg Infect Dis.

38. **Amarillas L**, **Rubí-Rangel L**, **Chaidez C**, **González-Robles A**, **Lightbourn-Rojas L**, **León-Félix J**. 2017. Isolation and characterization of phiLLS, a novel phage with potential biocontrol agent against multidrug-resistant Escherichia coli. Front Microbiol **8**:1–18.

39. **Neupane M**, **Abu-Ali GS**, **Mitra A**, **Lacher DW**, **Manning SD**, **Riordan JT**. 2011. Shiga toxin 2 overexpression in Escherichia coli O157:H7 strains associated with severe human disease. Microb Pathog **51**:466–470.

40. **Abu-Ali GS**, **Ouellette LM**, **Henderson ST**, **Whittam TS**, **Manning SD**. 2010. Differences in adherence and virulence gene expression between two outbreak strains of enterohaemorrhagic Escherichia coli O157:H7. Microbiology **156**:408–419.

41. **Schmittgen TD**, **Livak KJ**. 2008. Analyzing real-time PCR data by the comparative CT method. Nat Protoc **3**:1101–1108.

42. **Fernandes MA**, **Verstraete SG**, **Phan TG**, **Deng X**, **Stekol E**, **LaMere B**, **Lynch S V.**, **Heyman MB**, **Delwart E**. 2019. Enteric Virome and Bacterial Microbiota in Children With Ulcerative Colitis and Crohn Disease. J Pediatr Gastroenterol Nutr **68**:30–36.

43. **Reyes A**, **Blanton L V.**, **Cao S**, **Zhao G**, **Manary M**, **Trehan I**, **Smith MI**, **Wang D**, **Virgin HW**, **Rohwer F**, **Gordon JI**. 2015. Gut DNA viromes of Malawian twins discordant for severe acute malnutrition. Proc Natl Acad Sci U S A **112**:11941–11946.

44. **Dinakaran V**, **Rathinavel A**, **Pushpanathan M**, **Sivakumar R**, **Gunasekaran P**, **Rajendhran J**. 2014. Elevated levels of circulating DNA in cardiovascular disease patients: metagenomic profiling of microbiome in the circulation. PLoS One **9**:e105221.

45. **Gootenberg DB**, **Paer JM**, **Luevano J-M**, **Kwon DS**. 2016. HIV-associated changes in the enteric microbial community. Curr Opin Infect Dis **0**:1.

46. **Phan TG**, **Vo NP**, **Bonkoungou IJO**, **Kapoor A**, **Barro N**, **O'Ryan M**, **Kapusinszky B**, **Wang C**, **Delwart E**. 2012. Acute Diarrhea in West African Children: Diverse Enteric Viruses and a Novel Parvovirus Genus. J Virol **86**:11024–11030.

47. **Ngoi CN**, **Siqueira J**, **Li L**, **Deng X**, **Mugo P**, **Graham SM**, **Price MA**, **Sanders EJ**, **Delwart E**. 2016. The plasma virome of febrile adult kenyans shows frequent parvovirus B19 infections and a novel arbovirus (Kadipiro virus). J Gen Virol **97**:3359–3367.

48. **Young JC**, **Chehoud C**, **Bittinger K**, **Bailey A**, **Diamond JM**, **Cantu E**, **Haas AR**, **Abbas A**, **Frye L**, **Christie JD**, **Bushman FD**, **Collman RG**. 2015. Viral metagenomics reveal blooms of anelloviruses in the respiratory tract of lung transplant recipients. Am J Transplant **15**:200–209.

49. **Bernardin F**, **Operskalski E**, **Busch M**, **Delwart E**. 2010. Transfusion transmission of highly prevalent commensal human viruses. Transfusion.

50. **Pifferi M**, **Maggi F**, **Caramella D**, **De Marco E**, **Andreoli E**, **Meschi S**, **MacChia P**, **Bendinelli M**, **Boner AL**. 2006. High torquetenovirus loads are correlated with bronchiectasis and peripheral airflow limitation in children. Pediatr Infect Dis J.

51. **Wang J**, **Zheng J**, **Shi W**, **Du N**, **Xu X**, **Zhang Y**, **Ji P**, **Zhang F**, **Jia Z**, **Wang Y**, **Zheng Z**, **Zhang H**, **Zhao F**. 2018. Dysbiosis of maternal and neonatal microbiota associated with gestational diabetes mellitus. Gut.

52. **Monaco CL**, **Gootenberg DB**, **Zhao G**, **Handley SA**, **Ghebremichael MS**, **Lim ES**, **Lankowski A**, **Baldridge MT**, **Wilen CB**, **Flagg M**, **Norman JM**, **Keller BC**, **Lu??vano JM**, **Wang D**, **Boum Y**, **Martin JN**, **Hunt PW**, **Bangsberg DR**, **Siedner MJ**, **Kwon DS**, **Virgin HW**. 2016. Altered Virome and Bacterial Microbiome in Human

Immunodeficiency Virus-Associated Acquired Immunodeficiency Syndrome. Cell Host Microbe **19**:311–322.

53. **Dutilh BE**, **Cassman N**, **McNair K**, **Sanchez SE**, **Silva GGZ**, **Boling L**, **Barr JJ**, **Speth DR**, **Seguritan V**, **Aziz RK**, **Felts B**, **Dinsdale EA**, **Mokili JL**, **Edwards RA**. 2014. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. Nat Commun.

54. **Shkoporov AN**, **Khokhlova E V.**, **Fitzgerald CB**, **Stockdale SR**, **Draper LA**, **Ross RP**, **Hill C**. 2018. ΦCrAss001 represents the most abundant bacteriophage family in the human gut and infects Bacteroides intestinalis. Nat Commun **9**:1–8.

55. **Krishnamurthy SR**, **Wang D**. 2016. Origins and challenges of viral dark matter. Virus Res.

56. **Kasman LM**. 2005. Barriers to coliphage infection of commensal intestinal flora of laboratory mice. Virol J **2**:34.

57. **Meader E**, **Mayer MJ**, **Gasson MJ**, **Steverding D**, **Carding SR**, **Narbad A**. 2010. Bacteriophage treatment significantly reduces viable Clostridium difficile and prevents toxin production in an in vitro model system. Anaerobe **16**:549–554.

58. **Wendel BM**, **Cole JM**, **Courcelle CT**, **Courcelle J**. 2017. SbcC-SbcD and ExoI process convergent forks to complete chromosome replication. Proc Natl Acad Sci U S A.

59. **Lloyd RG**, **Buckman C**. 1985. Identification and genetic analysis of sbcC mutations in commonly used recBC sbcB strains of Escherichia coli K-12. J Bacteriol.

60. **Viazis S**. 2010. Control of Enterohemorrhagic Escherichia coli using bacteriophages. Stress Int J Biol Stress.

61. **Amarillas L**, **Chaidez C**, **González-Robles A**, **León-Félix J**. 2016. Complete genome sequence of new bacteriophage phiE142, which causes simultaneously lysis of multidrug-resistant Escherichia coli O157:H7 and Salmonella enterica. Stand Genomic Sci **11**:89.

62. **Goodridge L**, **Gallaccio A**, **Griffiths MW**. 2003. Morphological, host range, and genetic characterization of two coliphages. Appl Environ Microbiol **69**:5364–5371.

63. **Lenski RE**. 1988. Dynamics of Interactions between Bacterial and Virulent Bacteriophage. Adcances Microb Ecol **10**:1–44.

64. **Flynn GO**, **Ross RP**, **Fitzgerald GF**, **Coffey A**. 2004. Evaluation of a Cocktail of Three Bacteriophages for Biocontrol of Escherichia coli O157 : H7 Evaluation of a Cocktail of Three Bacteriophages for Biocontrol of Escherichia coli O157 : H7. Appl Environ Microbiol **70**:3417–3424.

65. **Tanji Y**, **Shimada T**, **Fukudomi H**, **Miyanaga K**, **Nakai Y**, **Unno H**. 2005. Therapeutic use of phage cocktail for controlling Escherichia coli O157:H7 in gastrointestinal tract of mice. J Biosci Bioeng **100**:280–7.

66. **Ross RP**, **Fitzgerald GF**, **Coffey A**. 2004. Evaluation of a Cocktail of Three Bacteriophages for Biocontrol of. Society **70**:3417–3424.

67. **Amor K**, **Heinrichs DE**, **Frirdich E**, **Ziebell K**, **Johnson RP**, **Whitfield C**. 2000. Distribution of core oligosaccharide types in lipopolysaccharides from Escherichia coli. Infect Immun.

68. **Sheng H**, **Knecht HJ**, **Kudva IT**, **Hovde CJ**. 2006. Application of bacteriophages to control intestinal Escherichia coli O157:H7 levels in ruminants. Appl Environ Microbiol **72**:5359–5366.

69. **Mizoguchi K**, **Morita M**, **Fischer CR**, **Yoichi M**, **Tanji Y**, **Unno H**. 2003. Coevolution of bacteriophage PP01 and Escherichia coli O157 : H7 in continuous culture. Appl Environ Microbiol **69**:170–176.

70. **Proctor RA**, **von Eiff C**, **Kahl BC**, **Becker K**, **McNamara P**, **Herrmann M**, **Peters G**. 2006. Small colony variants: a pathogenic form of bacteria that facilitates persistent and recurrent infections. NatRevMicrobiol **4**:295–305.

71. **Mizoguchi K**, **Morita M**, **Fischer CR**, **Yoichi M**, **Tanji Y**, **Unno H**. 2013. Using experimental evolution to optimize combination therapy : phages + antibiotics vs Pseudomonas aeruginosa Supervised by Michael Hochberg. Appl Environ Microbiol **69**:2012–2013.

72. **Scheiring J**, **Andreoli SP**, **Zimmerhackl LB**. 2008. Treatment and outcome of Shiga-toxin-associated hemolytic uremic syndrome (HUS). Pediatr Nephrol.

73. **Trachtman H**, **Austin C**, **Lewinski M**, **Stahl R a. K**. 2012. Renal and neurological involvement in typical Shiga toxin-associated HUS. Nat Rev Nephrol **8**:658–669.

74. **Mühldorfer I**, **Hacker J**, **Keusch GT**, **Acheson DW**, **Tschäpe H**, **Kane A V.**, **Ritter A**, **Ölschläger T**, **Donohue-Rolfe A**. 1996. Regulation of the Shiga-like toxin II operon in Escherichia coli. Infect Immun **64**:495–502.

75. **Acheson DWK**, **Reidl J**, **Zhang X**, **Keusch GT**, **Mekalanos JJ**, **Waldor MK**. 1998. In vivo transduction with shiga toxin 1-encoding phage. Infect Immun.

76. **Schmidt H**, **Bielaszewska M**, **Karch H**. 1999. Transduction of enteric Escherichia coli isolates with a derivative of Shiga toxin 2-encoding bacteriophage ??3538 isolated from Escherichia coli O157:H7. Appl Environ Microbiol.

77. **James CE**, **Stanley KN**, **Allison HE**, **Flint HJ**, **Stewart CS**, **Sharp RJ**, **Saunders JR**, **Mccarthy AJ**. 2001. Lytic and Lysogenic Infection of Diverse Escherichia coli and Shigella Strains with a Verocytotoxigenic Bacteriophage. Appl Environ Microbiol.

78. **Tyler JS**, **Mills MJ**, **Friedman DI**. 2004. The operator and early promoter region of the Shiga toxin type 2-encoding bacteriophage 933W and control of toxin expression. J Bacteriol.

79. **Koudelka AP**, **Hufnagel LA**, **Koudelka GB**. 2004. Purification and characterization of

the repressor of the shiga toxin-encoding bacteriophage 933W: DNA binding, gene regulation, and autocleavage. J Bacteriol.

80. **Wagner PL**, **Neely MN**, **Zhang X**, **Acheson DWK**, **Waldor MK**, **Friedman DI**. 2001. Role for a phage promoter in Shiga toxin 2 expression from a pathogenic Escherichia coli strain. J Bacteriol.

81. **Livny J**, **Friedman DI**. 2004. Characterizing spontaneous induction of Stx encoding phages using a selectable reporter system. Mol Microbiol.

82. **Kimmitt PT**, **Harwood CR**, **Barer MR**. 2000. Toxin gene expression by shiga toxin-producing Escherichia coli: the role of antibiotics and the bacterial SOS response. Emerg Infect Dis **6**:458–65.

83. **Fuchs S**, **Mühldorfer I**, **Donohue-Rolfe A**, **Kerényi M**, **Emödy L**, **Alexiev R**, **Nenkov P**, **Hacker J**. 1999. Influence of RecA on in vivo virulence and Shiga toxin 2 production in Escherichia coli pathogens. Microb Pathog.

84. **Bloch S**, **Nejman-Faleńczyk B**, **Topka G**, **Dydecka A**, **Licznerska K**, **Narajczyk M**, **Necel A**, **Węgrzyn A**, **Węgrzyn G**. 2015. UV-sensitivity of shiga toxin-converting bacteriophage virions Φ24<inf>B</inf>, 933W, P22, P27 and P32. Toxins (Basel).

85. **Wagner PL**, **Acheson DWK**, **Waldor MK**. 2001. Human neutrophils and their products induce Shiga toxin production by enterohemorrhagic Escherichia coli. Infect Immun.

86. **Imamovic L**, **Muniesa M**. 2012. Characterizing RecA-independent induction of Shiga toxin2-encoding phages by EDTA treatment. PLoS One.

87. **McGannon CM**, **Fuller CA**, **Weiss AA**. 2010. Different classes of antibiotics differentially influence shiga toxin production. Antimicrob Agents Chemother **54**:3790–3798.

88. **de Sablet T**, **Bertin Y**, **Vareille M**, **Girardeau JP**, **Garrivier A**, **Gobert AP**, **Martin C**. 2008. Differential expression of stx2 variants in Shiga toxin-producing Escherichia coli belonging to seropathotypes A and C. Microbiology **154**:176–186.

89. **Campoy S**, **Hervàs A**, **Busquets N**, **Erill I**, **Teixidó L**, **Barbé J**. 2006. Induction of the SOS response by bacteriophage lytic development in Salmonella enterica. Virology.

90. **Smith HO**, **Levine M**. 1965. The synthesis of phage and host dna in the establishment of lysogeny. Virology.

91. **Woodworth-Gutai M**, **Israel V**, **Levine M**. 1972. New deoxyribonuclease activity after bacteriophage P22 infection. J Virol.

92. **Schmieger H**, **Buch U**. 1975. Appearance of transducing particles and the fate of host DNA after infection of Salmonella typhimurium with P22-mutants with increased transducing ability (HT-mutants). MGG Mol Gen Genet.

93. **Chevallereau A**, **Blasdel BG**, **De Smet J**, **Monot M**, **Zimmermann M**, **Kogadeeva M**,

Sauer U, **Jorth P**, **Whiteley M**, **Debarbieux L**, **Lavigne R**. 2016. Next-Generation "-omics" Approaches Reveal a Massive Alteration of Host RNA Metabolism during Bacteriophage Infection of Pseudomonas aeruginosa. PLoS Genet.

94. **Veses-Garcia M**, **Liu X**, **Rigden DJ**, **Kenny JG**, **McCarthy AJ**, **Allison HE**. 2015. Transcriptomic analysis of shiga-toxigenic bacteriophage carriage reveals a profound regulatory effect on acid resistance in Escherichia coli. Appl Environ Microbiol **81**:8118–8125.

95. **Comeau AM**, **Tétart F**, **Trojet SN**, **Prère MF**, **Krisch HM**. 2007. Phage-antibiotic synergy (PAS): β-lactam and quinolone antibiotics stimulate virulent phage growth. PLoS One **2**:8–11.

96. **Kirby AE**. 2012. Synergistic Action of Gentamicin and Bacteriophage in a Continuous Culture Population of Staphylococcus aureus. PLoS One **7**.

97. **Coulter LB**, **McLean RJC**, **Rohde RE**, **Aron GM**. 2014. Effect of bacteriophage infection in combination with tobramycin on the emergence of resistance in Escherichia coli and Pseudomonas aeruginosa biofilms. Viruses **6**:3778–3786.

98. **Kim M**, **Jo Y**, **Hwang J**, **Hong W**, **Hong S**, **Park K**. 2018. Phage-Antibiotic Synergy via Delayed Lysis. Appl Environ Microbiol **84**:1–11.

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

Acute gastroenteritis is one of the most common illnesses associated with hospitalization globally (1). The number of acute cases of gastroenteritis annually is staggering; 2.3 billion cases of acute gastroenteritis and 1.3 million deaths occur each year worldwide (2). There are healthcare disparities based on geographic location. Developing countries have the most significant disease burden associated with acute gastroenteritis due in part to lack of infrastructure. While diarrheal illness accounts for 8% of all deaths in children under the age of five globally (3), one in eight deaths occur in children under the age of five (12.5%) in developing countries (4). In the United States, the number of annual cases ranges from 179 million (5) to 375 million (6), though many cases are not reported, given that some infections are self-limiting. Children are affected more severely by acute gastroenteritis in the United States, which contributes to 1.5 million office visits, 200,000 hospitalizations, and 300 deaths annually (7). Importantly, a subset of patients can have persistent long-term complications such as post-infectious irritable bowel syndrome with symptoms lasting up to 10 years (8, 9) or inflammatory bowel disease (IBD) (10). Mouse models have shown a potential mechanism in defining the movement from acute gastroenteritis to chronic conditions (11), which is driven primarily by host immunity due to changes in the microbiota from the infection (12). The microbiome is the genetic signature of the microbiota that inhabit a given environment. Although the function of a healthy intestinal microbiome has been elucidated, less is known about the impact of pathogen invasion. Defining the alterations in the human microbiome of the gastrointestinal (GI) tract due to acute bacterial gastroenteritis can aid in the development of prevention practices and in the identification of novel therapeutic targets that can be used to restore the microbiome to a healthy state.

Human DNA has been suggested to represent a contaminant of the gut microbiome (13). In our studies, intestinal microbial communities from patients with enteric infections (cases) had

a higher proportion of sequencing reads, on average, that mapped to human DNA sequences (15.2%) compared to healthy individuals (controls) (0.1%, Chapter 2) or patients' post-recovery (follow-ups) (0.1%, Chapter 3). This finding is consistent with studies on *Clostridium difficile* infections (14) and IBD (15) and colorectal cancer (16) patients, which have identified an increased quantity of human DNA in stool samples. This increase is likely due to the destruction of epithelial cells lining the GI tract. In enteric infections, tissue destruction is most likely a result of hemorrhagic colitis, which results in a release of nutrients such as carbon sources, vitamins, and minerals to the microbiota. This release of the cellular contents could provide the necessary nutrients to drive the observed dysbiosis in gastroenteritis. A future investigation into the metabolic profiles of the reads identified in this study is therefore warranted and could define critical metabolic pathways that are enriched during acute gastroenteritis infections.

Proteobacteria was the most differentially abundant phyla detected in the intestinal microbiomes of patients with gastroenteritis and was significantly higher than in healthy controls (17, 18). This finding is consistent with other studies showing that increased abundance of Proteobacteria is associated with inflammation and contribute to dysbiosis in gastroenteritis and other disease states such as HIV and IBS (18–22). Cases also had a higher abundance of Proteobacteria compared to the follow-up samples, suggesting that the dysbiosis can be corrected following recovery. It is important to note that the alterations in Proteobacteria populations have been shown to vary across pathogens, as was observed previously (17). Indeed, each case was caused by either *Salmonella* spp., *Shigella* spp., *Campylobacter jejuni*, or Shiga toxin-producing *Escherichia coli* (STEC), all of which belong to the Proteobacteria phylum. The abundance of *Escherichia*, in particular, was significantly increased among cases regardless of the infecting agent.

In contrast, the microbiome of healthy individuals, including uninfected controls and patients following recovery (follow-ups), had a higher abundance of Bacteroidetes and Firmicutes compared to cases, which have been observed (17, 18, 23). Previous studies have identified genera *Roseburia*, *Blauta,* and *Lachnospiraceae* to be more abundant in healthy people (17), which we have confirmed in our analysis using ANCOM. We further found that decreased relative abundance of *Roseburia* was associated with more severe illness as it had decreased abundance in Cluster 2. Members of *Roseburia* represent a group of butyrate producers (24), which were suggested to dampens the immune response through nuclear kappa B and improves colitis in mouse models (25). Hence, decreasing the abundance of butyrate-producing microbes could increase the local immune response through lowered butyrate production.

The bacterial component of the intestinal microbiome has been well characterized, however, less is known about the virome, or the collection of resident viruses, particularly during acute gastroenteritis infections. Prior studies have examined the virome with multiple displacement amplification (26), direct isolation of viruses with sequencing (27, 28), and identification in metagenomes (29). Through these studies it has become apparent that viral databases are incomplete (30) with many isolated viral particles not aligning to known sequences (27, 31–34), assemblies of reads from metagenomes of isolated viruses have yielded less than 2% taxonomically annotated (35); this is in stark contrast to bacterial databases that can achieve greater than 90% identification of the diversity in the sequences down to the species level (36). This lack of annotation is referred to as “viral dark matter” and is due to the relatively small size of known viruses (30). Many of the entries in viral databases are predominately filled with *Escherichia* bacteriophage. Cross-assembly has been utilized to find a highly abundant bacteriophage within metagenomic datasets (37) and could be used on the sequences reported in

Chapters 2 and 3 in future analyses to detect and characterize unknown viruses. Newer, more encompassing viral databases such as the reference viral database (38), need to be utilized to improve the annotation rate of the sequences in virome studies.

A recent study published in 2018 (39), recommended an investigation into the viruses present during episodes of gastroenteritis. Indeed, these viral populations were explored in Chapters 2 and 3, utilizing a kmer-based sequence annotation approach (40). Caudovirales, a significant family of bacteriophage, was increased in abundance among gastroenteritis patients compared to the uninfected controls and patients at follow-up or recovery. Previous studies have identified increased Caudovirales abundance and diversity within different patient populations. Piggyback-the-winner (41) states that bacteriophage abundance increases in response to an increase in the host population, which the bacteriophage utilizes for replication; this is likely the observation here. Indeed, the primary genera of bacteriophage found to be elevated in cases are *Nona33virus*, *P2virus*, *P1virus*, which infect Enterobacteriaceae hosts.

Hierarchical clustering identified four distinct clusters of microbial profiles that differed by study group. Notably, Cluster 2 was composed of patients that had a more severe illness relative to other cases (Chapter 2). Differential abundance analysis identified 92 genera that varied in patients with microbiome profiles belonging to Cluster 2 when compared to the post-recovery profiles (Chapter 3); only 82 genera were differentially abundant when compared to controls (Chapter 2). This difference could be due to individual variation in the microbiome, sample size differences between the two studies, or the follow-up were different from the healthy populations. We identified three genera (*Alistipes*, *Sutterella*, *Odoribacter*) that were lower in abundance among the follow-up samples compared to the control samples. The role of these three bacterial populations is not fully known. *Alistipes* have been correlated positively with

health (42), and *Odoribacter* produces butyrate and could be significant in regulating inflammation (43), and *Sutterella* is a commensal that might aid in immune regulation (44). These microbial populations could be investigated for their role in intestinal health in future studies as they may be able to facilitate faster recovery times.

Use of logistic regression identified different microbial populations to be important for enteric infections among cases relative to otherwise healthy individuals; differences were observed between the study groups described in Chapters 2 (case vs. control) and Chapter 3 (case vs. follow-up). Cases with more severe infections and microbial communities belonging to Cluster 2 were more likely to have an increased abundance of *Actinobacteria*, *Orthopoxvirus*, *Salmonella,* and *Serratia* relative to all other samples. The identification of these four taxa is biologically plausible as both *Actinobacter* (45), and *Orthopoxvirus* (46–48) have been shown to interact with the immune system. However, Chapter 3 identified that three genera *Shigella, Enterobacter*, and *Pantoea* were predictors of microbial communities belonging to Cluster 2, though the controls were not included in this analysis. *Enterobacter* and *Pantoea* were found to be differentially abundant in Cluster 2 compared to the rest of the samples, which suggests that these microbial populations could serve as indicator organisms in patients to enhance detection of those with more severe clinical outcomes. The different findings between the chapters are due to the comparison of different samples as hierarchical clustering is dependent on the type and number of samples used in any given analysis.

Further evaluation and validation of the results are needed through additional studies. A meta-analysis of the gastroenteritis studies available should be undertaken to determine the complete picture. The meta-analysis should (if possible) directly combine the sequencing data that is available for each study, as was previously done with the microbiome and diet (49).

These findings can then be evaluated in mouse models. Previous studies in mice have identified changes in the microbiome due to an exogenous challenge (11, 50). *Citrobacter rodentium* (the mouse equivalent to *Salmonella*), has been used to mimic gastroenteritis in mice (11) and can be used as a model for future studies. Mouse models, however, have limitations such as varying physiology, anatomy, diet, genetics, housing, and immune responses (51), which can impact interpretations regarding the human microbiome. Indeed, the microbiomes are distinctly different in mice and humans. Humans have been shown to have a greater abundance of *Prevotella*, *Faecalibacterium*, and *Ruminoccus,* whereas the dominant mouse gut microbiota consists of *Lactobacillus*, *Alistipes*, and *Turicibacter* (51). Clostridium, Bacteroides and Blautia were found to be shared between humans and mice. Despite these differences, studies in mice have been successfully utilized to study inflammatory triggers during colitis (12). It may be possible to directly study the genera identified in Chapters 2 and 3 within mouse models to examine their effects. This approach could provide insight into the pathogenesis of gastroenteritis within mouse models and identify distinct ecological niches making the findings here more generalizable.

Because the microbiome has been shown to vary across individuals in different geographic locations (52), this study consisted mainly of a Caucasian population that resided in Michigan. Repeating this study at a different site either within the United States or elsewhere could enhance understanding of the microbiome in a different setting while evaluating the impact of factors such as diet, infectious agents, etc., which could differentially impact the microbial populations. Identification of similar factors will also improve the generalizability of the results observed herein. It is also important to note that the microbiome will differ at different locations along the GI tract (53), and hence, the resident populations identified in this study will not necessarily match those that occur during gastroenteritis in different locations.

Since a prior study showed that bacteriophages adhere to mucosa and can prevent pathogen infection of eukaryotic cells (54), we expected to observe an enrichment of pathogen-specific bacteriophages in the presence of a given pathogen. Indeed, we observed that isolated viral-like particles (VLPs) were six times more likely to lyse pathogenic strains than commensal *E. coli* strains. These findings suggest that bacteriophage capable of lysing pathogens are more common in intestinal microbial communities. Novel bacteriophages using *E. coli* as hosts have been isolated from fecal samples previously (55, 56). Notably, we isolated three novel bacteriophages, which were classified genetically as two lysogenic bacteriophages and one lytic bacteriophage (Chapter 4). Homologous sequences for the lysogenic bacteriophages were detected as prophages in 23 additional *E. coli* strains, while both bacteriophages carried the exonuclease *sbcC* that is essential for DNA replication and repair (57). Additional studies are needed to determine whether there is a benefit to the bacterial host that possesses an extra copy of the DNA repair gene; we hypothesize that the additional copy of *sbcC* results in fewer mutations and results in a more stable genome. The lytic bacteriophage, PHG001, demonstrated high selectivity and virulence towards multiple *E. coli* O157:H7 strains, though bacterial resistance towards PHG001 was observed by 24 hours. The resistant phenotype has a rough appearance (58, 59), which could be due to alterations in the cell wall or O-antigen given the high specificity of PHG001 to specific O-types. PHG001 also reduced expression of the Shiga toxin gene, *stx2c*, after infection of *E. coli* O157:H7 for three hours. Collectively, these results add knowledge to bacteriophage genomics and bacteriophage-host relationships and highlight the importance of better defining relationships within microbial communities.

In future studies, a massively parallel screening approach could be utilized to investigate the role individual microbes play within the overall microbial community (60). This approach

would allow for the building of synthetic microbial communities and an assessment of the interactions and effects of different microbes in different conditions (60). Artificial microbiomes can be built that foster the growth of bacteria like *Odoribacter,* identified to be important for a recovered microbiome in our study. This type of analysis will allow us to understand how the microbiome recovers mechanistically from a perturbation such as illness. Translation of these studies could lead to improvement in therapeutics including fecal microbiota transplants (FMT).

Fecal microbiota transplants (61) (FMT) involve the transfer of stool containing microbiota from healthy donors to patients to restore the microbiome to a healthy state(62). Cure rates of *Clostridium difficile* with the use of FMTs have been reported as high as 90% (63), and there are current FMT treatment investigations on metabolic syndrome (64), autism spectrum disorder (61), and IBD (65). Low counts of *Faecalibacterium* have been observed in IBD patients (66), which FMT attempts to restore via transfer from a healthy donor. Additional studies could investigate the impact of transferring genera identified here, namely *Roseburia*, *Alistipes,* and *Odoribacter*. The success of FMT has been linked to species richness, or the number of microbes, in the microbiome from the donor (67). Virus particles are transferred during an FMT (68), and higher numbers of unique bacteriophages in the donor are strongly correlated with the success of an FMT (69). Additionally, the use of bacteriophage could represent a targeted approach to prevent the overgrowth of *Escherichia* during gastroenteritis. Such an approach could increase recovery times and decrease the chronic disease burden as a prior colitis study found that blocking the overgrowth of *Escherichia* during gastroenteritis could improve health in mice (11).

In summary, we have comprehensively examined the microbiome in patients with acute bacterial gastroenteritis for comparison to healthy uninfected individuals and a subset of the

same patients post-recovery. Additionally, we have isolated and characterized three novel bacteriophages, and have examined the function of one lytic phage in multiple bacterial hosts. Collectively, these findings have improved our knowledge of acute bacterial gastroenteritis with the use of bioinformatics, and have identified specific microbiome profiles to be associated with more severe infections. These data provide insight into new prevention strategies and novel therapies to potentially facilitate treatment and recovery from acute bacterial gastroenteritis.

REFERENCES

REFERENCES

1. **World Health Organization**. 2017. Diarrhoeal disease. Clin Med.

2. **Vos T, Allen C, Arora M, Barber RM, Brown A, Carter A, Casey DC, Charlson FJ, Chen AZ, Coggeshall M, Cornaby L, Dandona L, Dicker DJ, Dilegge T**,. 2016. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 388:1545–1602.

3. **UNICEF**. Diarrhoeal Disease: Current Status + Progress.

4. **Kotloff KL**. 2017. The Burden and Etiology of Diarrheal Illness in Developing Countries. Pediatr Clin North Am 64:799–814.

5. **Hall AJ, Rosenthal M, Gregoricus N, Greene SA, Ferguson J, Henao OL, Vinjé J, Lopman BA, Parashar UD, Widdowson MA**. 2011. Incidence of acute gastroenteritis and role of norovirus, Georgia, USA, 2004-2005. Emerg Infect Dis 17:1381–1388.

6. **Herikstad. H, Yang S, Van Gilder TJ, Vugia D, Hadler J, Blake P, Deneen V, Shiferaw B, Angulo FJ**. 2002. A population-based estimate of the burden of diarrhoeal illness in the United States: FoodNet, 1996-7. Epidemiol Infect 129:9–17.

7. **American Academy of Family Physicians., Hartman S, Brown E, Loomis E, Russell HA**. 2019. Gastroenteritis in Children. Am Fam Physician 99:159–165.

8. **Schwille-Kiuntke J, Enck P, Zendler C, Krieg M, Polster A V., Klosterhalfen S, Autenrieth IB, Zipfel S, Frick JS**. 2011. Postinfectious irritable bowel syndrome: Follow-up of a patient cohort of confirmed cases of bacterial infection with Salmonella or Campylobacter. Neurogastroenterol Motil.

9. **Thabane M, Simunovic M, Akhtar-Danesh N, Garg AX, Clark WF, Collins SM, Salvadori M, Marshall JK**. 2010. An outbreak of acute bacterial gastroenteritis is associated with an increased incidence of irritable bowel syndrome in children. Am J Gastroenterol.

10. **Rodríguez LAG, Ruigómez A, Panés J**. 2006. Acute Gastroenteritis Is Followed by an Increased Risk of Inflammatory Bowel Disease. Gastroenterology 130:1588–1594.

11. **Small CL, Xing L, McPhee JB, Law HT, Coombes BK**. 2016. Acute Infectious Gastroenteritis Potentiates a Crohn's Disease Pathobiont to Fuel Ongoing Inflammation in the Post-Infectious Period. PLoS Pathog 12:1–20.

12. **Lupp C, Robertson ML, Wickham ME, Sekirov I, Champion OL, Gaynor EC, Finlay BB**. 2007. Host-Mediated Inflammation Disrupts the Intestinal Microbiota and Promotes the Overgrowth of Enterobacteriaceae. Cell Host Microbe 2:119–129.

13. **Schmieder R, Edwards R**. 2011. Fast identification and removal of sequence contamination from genomic and metagenomic datasets. PLoS One.

14. **Vincent C, Mehrotra S, Loo VG, Dewar K, Manges AR**. 2015. Excretion of Host DNA in Feces Is Associated with Risk of Clostridium difficile Infection. J Immunol Res.

15. **Lewis JD, Chen EZ, Baldassano RN, Otley AR, Griffiths AM, Lee D, Bittinger K, Bailey A, Friedman ES, Hoffmann C, Albenberg L, Sinha R, Compher C, Gilroy E, Nessel L, Grant A, Chehoud C, Li H, Wu GD, Bushman FD**. 2015. Inflammation, Antibiotics, and Diet as Environmental Stressors of the Gut Microbiome in Pediatric Crohn's Disease. Cell Host Microbe.

16. **Klaassen CHW, Jeunink MAF, Prinsen CFM, Ruers TJM, Tan ACITL, Strobbe LJA, Thunnissen FBJM**. 2003. Quantification of human DNA in feces as a diagnostic test for the presence of colorectal cancer. Clin Chem.

17. **Singh P, Teal TK, Marsh TL, Tiedje JM, Mosci R, Jernigan K, Zell A, Newton DW, Salimnia H, Lephart P, Sundin D, Khalife W, Britton RA, Rudrik JT, Manning SD**. 2015. Intestinal microbial communities associated with acute enteric infections and disease recovery. Microbiome 3:45.

18. **Castaño-Rodríguez N, Underwood AP, Merif J, Riordan SM, Rawlinson WD, Mitchell HM, Kaakoush NO**. 2018. Gut microbiome analysis identifies potential etiological factors in acute gastroenteritis. Infect Immun 86:1–13.

19. **Shin NR, Whon TW, Bae JW**. 2015. Proteobacteria: Microbial signature of dysbiosis in gut microbiota. Trends Biotechnol.

20. **Monaco CL, Gootenberg DB, Zhao G, Handley SA, Ghebremichael MS, Lim ES, Lankowski A, Baldridge MT, Wilen CB, Flagg M, Norman JM, Keller BC, Lu??vano JM, Wang D, Boum Y, Martin JN, Hunt PW, Bangsberg DR, Siedner MJ, Kwon DS, Virgin HW**. 2016. Altered Virome and Bacterial Microbiome in Human Immunodeficiency Virus-Associated Acquired Immunodeficiency Syndrome. Cell Host Microbe 19:311–322.

21. **Nelson AM, Walk ST, Taube S, Taniuchi M, Houpt ER, Wobus CE, Young VB**. 2012. Disruption of the Human Gut Microbiota following Norovirus Infection. PLoS One 7.

22. **Saulnier DM, Riehle K, Mistretta TA, Diaz MA, Mandal D, Raza S, Weidler EM, Qin X, Coarfa C, Milosavljevic A, Petrosino JF, Highlander S, Gibbs R, Lynch S V., Shulman RJ, Versalovic J**. 2011. Gastrointestinal microbiome signatures of pediatric patients with irritable bowel syndrome. Gastroenterology.

23. **Youmans BP, Ajami NJ, Jiang Z-D, Campbell F, Wadsworth WD, Petrosino JF, DuPont HL, Highlander SK**. 2015. Characterization of the human gut microbiome during travelers' diarrhea. Gut Microbes 6:110–119.

24. **Duncan SH, Hold GL, Barcenilla A, Stewart CS, Flint HJ**. 2002. Roseburia intestinalis sp. nov., a novel saccharolytic, butyrate-producing bacterium from human faeces. Int J Syst Evol Microbiol.

25. **Segain JP, Galmiche JP, Raingeard De La Blétière D, Bourreille A, Leray V, Gervois N, Rosales C, Ferrier L, Bonnet C, Blottière HM**. 2000. Butyrate inhibits inflammatory responses through NFκB inhibition: Implications for Crohn's disease. Gut.

26. **Zhang T, Breitbart M, Lee WH, Run JQ, Wei CL, Soh SWL, Hibberd ML, Liu ET,**

Rohwer F, Ruan Y. 2006. RNA viral community in human feces: Prevalence of plant pathogenic viruses. PLoS Biol 4:0108–0118.

27. **Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, Rohwer F, Gordon JI**. 2010. Viruses in the faecal microbiota of monozygotic twins and their mothers. Nature 466:334–338.

28. **Breitbart M, Hewson I, Felts B, Mahaffy JM, Nulton J, Salamon P, Rohwer F**. 2003. Metagenomic analyses of an uncultured viral community from human feces. J Bacteriol 185:6220–6223.

29. **Roux S, Enault F, Hurwitz BL, Sullivan MB**. 2015. VirSorter: Mining viral signal from microbial genomic data. PeerJ 2015:1–20.

30. **Krishnamurthy SR, Wang D**. 2016. Origins and challenges of viral dark matter. Virus Res.

31. **Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, Wu GD, Lewis JD, Bushman FD**. 2011. The human gut virome: Inter-individual variation and dynamic response to diet. Genome Res 21:1616–1625.

32. **Handley SA, Desai C, Zhao G, Droit L, Monaco CL, Schroeder AC, Nkolola JP, Norman ME, Miller AD**, Wang D, Barouch DH, Virgin HW. 2016. SIV Infection-Mediated Changes in Gastrointestinal Bacterial Microbiome and Virome Are Associated with Immunodeficiency and Prevented by Vaccination. Cell Host Microbe 19:323–335.

33. **Reyes A, Blanton L V., Cao S, Zhao G, Manary M, Trehan I, Smith MI, Wang D, Virgin HW, Rohwer F, Gordon JI**. 2015. Gut DNA viromes of Malawian twins discordant for severe acute malnutrition. Proc Natl Acad Sci U S A 112:11941–11946.

34. **Minot S, Bryson A**. 2013. Rapid evolution of the human gut virome. Proc … 110:12450–12455.

35. **Shkoporov AN, Ryan FJ, Draper LA, Forde A, Stockdale SR, Daly KM, McDonnell SA, Nolan JA, Sutton TDS, Dalmasso M, McCann A, Ross RP, Hill C**. 2018. Reproducible protocols for metagenomic analysis of human faecal phageomes. Microbiome 6:68.

36. **Browne HP, Forster SC, Anonye BO, Kumar N, Neville BA, Stares MD, Goulding D, Lawley TD**. 2016. Culturing of "unculturable" human microbiota reveals novel taxa and extensive sporulation. Nature.

37. **Dutilh BE, Cassman N, McNair K, Sanchez SE, Silva GGZ, Boling L, Barr JJ, Speth DR, Seguritan V, Aziz RK, Felts B, Dinsdale EA, Mokili JL, Edwards RA**. 2014. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. Nat Commun.

38. **Goodacre N, Aljanahi A, Nandakumar S, Mikailov M, Khan AS**. 2018. A Reference Viral Database (RVDB) To Enhance Bioinformatics Analysis of High-Throughput Sequencing for Novel Virus Detection. mSphere.

39. **Castaño-Rodríguez N, Underwood AP, Merif J, Riordan SM, Rawlinson WD, Mitchell**

HM, Kaakoush NO. 2018. Gut Microbiome Analysis Identifies Potential Etiological Factors in Acute Gastroenteritis. Infect Immun 86:e00060-18.

40. **Cottingham RW**. 2015. The DOE systems biology knowledgebase (KBase).

41. **Knowles B, Silveira CB, Bailey BA, Barott K, Cantu VA, Cobian-Guëmes AG, Coutinho FH, Dinsdale EA, Felts B, Furby KA, George EE, Green KT, Gregoracci GB, Haas AF, Haggerty JM, Hester ER, Hisakawa N, Kelly LW, Lim YW, Little M, Luque A, McDole-Somera T, McNair K, De Oliveira LS, Quistad SD, Robinett NL, Sala E, Salamon P, Sanchez SE, Sandin S, Silva GGZ, Smith J, Sullivan C, Thompson C, Vermeij MJA, Youle M, Young C, Zgliczynski B, Brainard R, Edwards RA, Nulton J, Thompson F, Rohwer F**. 2016. Lytic to temperate switching of viral communities. Nature.

42. **Santisteban MM, Qi Y, Zubcevic J, Kim S, Yang T, Shenoy V, Cole-Jeffrey CT, Lobaton GO, Stewart DC, Rubiano A, Simmons CS, Garcia-Pereira F, Johnson RD, Pepine CJ, Raizada MK**. 2017. Hypertension-Linked Pathophysiological Alterations in the Gut. Circ Res.

43. **Gomez-Arango LF, Barrett HL, McIntyre HD, Callaway LK, Morrison M, Dekker Nitert M**. 2016. Increased Systolic and Diastolic Blood Pressure is Associated with Altered Gut Microbiota Composition and Butyrate Production in Early Pregnancy. Hypertension.

44. **Hiippala K, Kainulainen V, Kalliomäki M, Arkkila P, Satokari R**. 2016. Mucosal prevalence and interactions with the epithelium indicate commensalism of Sutterella spp. Front Microbiol.

45. **Cekanaviciute E, Yoo BB, Runia TF, Debelius JW, Singh S, Nelson CA, Kanner R, Bencosme Y, Lee YK, Hauser SL, Crabtree-Hartman E, Sand IK, Gacias M, Zhu Y, Casaccia P, Cree BAC, Knight R, Mazmanian SK, Baranzini SE**. 2017. Gut bacteria from multiple sclerosis patients modulate human T cells and exacerbate symptoms in mouse models. Proc Natl Acad Sci U S A.

46. **Finlay BB, McFadden G**. 2006. Anti-immunology: Evasion of the host immune system by bacterial and viral pathogens. Cell.

47. **Alcami A**. 2003. Viral mimicry of cytokines, chemokines and their receptors. Nat Rev Immunol.

48. **De Cárcer DA, Hernáez B, Rastrojo A, Alcamí A**. 2017. Infection with diverse immune-modulating poxviruses elicits different compositional shifts in the mouse gut microbiome. PLoS One 12:1–9.

49. **Bisanz JE, Upadhyay V, Turnbaugh JA, Ly K, Turnbaugh PJ**. 2019. Meta-Analysis Reveals Reproducible Gut Microbiome Alterations in Response to a High-Fat Diet. Cell Host Microbe.

50. **Reyes A, Wu M, McNulty NP, Rohwer FL, Gordon JI**. 2013. Gnotobiotic mouse model of phage-bacterial host dynamics in the human gut. Proc Natl Acad Sci U S A 110:20236–20241.

51. **Nguyen TLA, Vieira-Silva S, Liston A, Raes J**. 2015. How informative is the mouse for

human gut microbiota research? DMM Dis Model Mech.

52. **Holtz LR, Cao S, Zhao G, Bauer IK, Denno DM, Klein EJ, Antonio M, Stine OC, Snelling TL, Kirkwood CD, Wang D**. 2014. Geographic variation in the eukaryotic virome of human diarrhea. Virology 468:556–564.

53. **Hillman ET, Lu H, Yao T, Nakatsu CH**. 2017. Microbial ecology along the gastrointestinal tract. Microbes Environ.

54. **Barr JJ, Auro R, Furlan M, Whiteson KL, Erb ML, Pogliano J, Stotland A, Wolkowicz R, Cutting AS, Doran KS, Salamon P, Youle M, Rohwer F**. 2013. Bacteriophage adhering to mucus provide a non-host-derived immunity. Proc Natl Acad Sci U S A 110:10771–6.

55. **Niu YD, McAllister TA, Nash JHE, Kropinski AM, Stanford K**. 2014. Four Escherichia coli O157:H7 phages: A new bacteriophage genus and taxonomic classification of T1-like phages. PLoS One.

56. **Morita M, Tanji Y, Mizoguchi K, Akitsu T, Kijima N, Unno H**. 2002. Characterization of a virulent bacteriophage specific for Escherichia coli O157:H7 and analysis of its cellular receptor and two tail fiber genes. FEMS Microbiol Lett 211:77–83.

57. **Wendel BM, Cole JM, Courcelle CT, Courcelle J**. 2017. SbcC-SbcD and ExoI process convergent forks to complete chromosome replication. Proc Natl Acad Sci U S A.

58. **Sheng H, Knecht HJ, Kudva IT, Hovde CJ**. 2006. Application of bacteriophages to control intestinal Escherichia coli O157:H7 levels in ruminants. Appl Environ Microbiol 72:5359–5366.

59. **Amor K, Heinrichs DE, Frirdich E, Ziebell K, Johnson RP, Whitfield C**. 2000. Distribution of core oligosaccharide types in lipopolysaccharides from Escherichia coli. Infect Immun.

60. **Kehe J, Kulesa A, Ortiz A, Ackerman CM, Thakku SG, Sellers D, Kuehn S, Gore J, Friedman J, Blainey PC**. 2019. Massively parallel screening of synthetic microbial communities. Proc Natl Acad Sci U S A.

61. **Kang D, Adams JB, Gregory AC, Borody T, Chittick L, Fasano A, Khoruts A, Geis E, Maldonado J, Mcdonough-means S, Pollard EL, Roux S, Sadowsky MJ, Lipson KS, Sullivan MB**. 2017. Microbiota Transfer Therapy alters gut ecosystem and improves gastrointestinal and autism symptoms : an open-label study. Microbiome 1–16.

62. **Broecker F, Klumpp J, Schuppler M, Russo G, Biedermann L, Hombach M, Rogler G, Moelling K**. 2016. Long-term changes of bacterial and viral compositions in the intestine of a recovered Clostridium difficile patient after fecal microbiota transplantation 1–13.

63. **Rohlke F, Stollman N**. 2012. Fecal microbiota transplantation in relapsing Clostridium difficile infection. Therap Adv Gastroenterol.

64. **Hartstra A V., Bouter KEC, Bäckhed F, Nieuwdorp M**. 2015. Insights into the role of the microbiome in obesity and type 2 diabetes. Diabetes Care.

65. **Sunkara T, Rawla P, Ofosu A, Gaduputi V**. 2018. Fecal microbiota transplant – a new frontier in inflammatory bowel disease. J Inflamm Res.

66. **Sokol H, Seksik P, Furet JP, Firmesse O, Nion-Larmurier I, Beaugerie L, Cosnes J, Corthier G, Marteau P, Doraé J**. 2009. Low counts of faecalibacterium prausnitzii in colitis microbiota. Inflamm Bowel Dis.

67. **Vermeire S, Joossens M, Verbeke K, Wang J, Machiels K, Sabino J, Ferrante M, Assche G Van, Rutgeerts P, Raes J**. 2016. Donor Species Richness Determines Faecal Microbiota Transplantation Success in Inflammatory Bowel Disease. J Crohn's Colitis 10:387–394.

68. **Chehoud C, Dryga A, Hwang Y, Nagy-Szakal D, Hollister EB, Luna RA, Versalovic J, Kellermayer R, Bushman FD**. 2016. Transfer of Viral Communities between Human Individuals during. MBio 7:1–8.

69. **Broecker F, Russo G, Klumpp J, Moelling K**. 2016. Stable core virome despite variable microbiome after fecal transfer. Gut Microbes 8:1–7.