BACTERIOPHAGE IN HOST ASSOCIATED MICROBIAL COMMUNITIES

EXAMINED WITH CONTINUOUS CULTURE SYSTEMS

by

Michael Stefan Dills

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Microbiology

MONTANA STATE UNIVERSITY
Bozeman, Montana

April 2018

ACKNOWLEDGMENTS

Many people have been involved with this project. In addition to my academic committee, Pilar Manrique has been instrumental in providing technical guidance with metagenomic analysis and computation. Beyond MSU, Doctors Jennifer Auchtung and Rob Britton, our collaborators at the Baylor School of Medicine, graciously offered time and expertise carrying out bioreactor experiments.

TABLE OF CONTENTS

TABLE OF CONTENTS CONTINUED

LIST OF TABLES

LIST OF FIGURES

Figure Page

GLOSSARY

Continuous Culture Systems: a semi-closed system in which select microbes are allowed to grow with constant influx of sterile media. Total volume is maintained by outflow of culture liquid and homogeneity is achieved through mixing.

Chemostat: an alternate term for continuous culture devices literally indicating the maintenance of a static chemical environment

Minibioreactor Arrays (MBRAs): the particular assembly of parallel chemostats developed at Baylor School of Medicine and used for the experimentation described in this thesis.

"-ome": application of this suffix to a biological classifier is meant to reference complete pool of all units within that classification. Phageome, virome, microbiome and mobilome will be used within this text. "-ome" has also been adopted to reference assemblages of biochemical species as in 'metabalome' and 'proteome'.

"-biome": as in 'microbiome', this refers to the complete microbial ecosystem of a given location including biological and molecular species as well as the physical environment containing them. Note that 'microbiome' is ambiguously defined and must be operationally interpreted.

"-biota": unlike '-ome' this modifier references the assembly of taxa within a classification. While 'microbiota' is the common form, semantic precision calls for equivalent terminology to the various –omes listed above.

Health: any one of a multitude of dynamic states in which no overt disease is present.

Stability: the ability of a dynamic community state to resist change (resistance) or recover from disruption (resilience).

Dysbiosis: a qualitative disruption of a microbiota's commensal and mutualistic capacity. This term is often operationally defined within the context of taxonomic or functional aspects of a community. Dysbiosis is not a requisite criteria in health or disease though it is implicated in the etiology, persistence and progression of certain diseases.

ABSTRACT

Mechanistic understanding of the role of extracellular and parasitic elements in host ecosystems is currently lacking. Extensive surveys have catalogued a large diversity of bacteriophage which associate differentially with definable host states. This work is an attempt to aid in the development of a coherent model for complex symbiosis within mammalian host ecosystems by investigating the role of bacteriophage in microbial community structure. It details an investigation of continuous culture systems as a platform to study bacteriophage within polymicrobial communities of the human GI tract. It then describes an experiment testing an extracellular community's ability to modulate bacterial community structure.

CHAPTER ONE

INTRODUCTION TO BACTERIOPHAGE IN HOST ASSOCIATED MICROBIAL COMMUNITIES AND RESEARCH HYPOTHESIS

Introduction

Host Associated Environments

Large multicellular organisms exist as functional ecosystems, harboring diverse symbiotic communities of bacteria, archaea, and invertebrate eukaryotes. Cellular assemblages, in turn, host a variety of parasitic elements including viruses, transposons, and other semi-autonomous mobile genetic and structural material.

These nested communities often form commensal and mutualistic relationships with their host by providing essential functions and creating a durable ecosystem foundation. However, in certain circumstances, this dynamic balance is disrupted or never formed, resulting in dysbiosis which can either be a response to or effecter of disease (1).

Human Associated Microbial Communities

Humans represent one of the most extensively characterized metazoan ecosystems thanks to broad initiatives like the Human Microbiome Project (HMP) and Metagenomics of the Human Intestinal Tract (MetaHIT) (2-4). Research has shown that physical features, or habitats, of the human host are deterministic in terms of microbial community structure and that intrapersonal variation of body-sites through time is relatively stable

while interpersonal comparisons reveal distinct taxonomic community profiles for individuals (5, 6). Host genetics, health status and environmental variables account for a great deal of interpersonal variation (3, 7, 8).

A majority of each human associated microbiota resides in the gastrointestinal tract (9). This community, often approximated by surveys of the distal colon or feces, is dominated by bacteria and hallmarked by low taxonomic diversity at the phyla level with high species and strain richness (10-14). Published estimates of microbial cell density range between $10^{11} – 10^{12}$ cells per gram of dry feces accounting for approximately 50 percent of total fecal biomass or colonic content (15-17). Aside from richness, community architecture in this environment is highly uneven with alternating subsets of species accounting for the majority of cells (18).

Certain variations of the gastrointestinal microbiota are associated with a number of host disease phenotypes (1, 19-22). These variations can be seen in the compositional and metabolic profiles of the microbiota. For instance, GI microbial communities from obese and lean individuals vary in their capacity to produce certain short chain fatty acids (23, 24). Obesity also correlates with increased Firmicutes to Bacteroidetes ratio (25-29). Inflammation and dysbiosis have been linked to increased Proteobacteria abundance in the GI tract (30). This is not an exhaustive account of work in the field of host associated microbiomes but a selection of some key studies in the area. In a conceptual framework, resistance and resilience constitute microbiota stability which, by extension relates to host phenotype and homeostasis. However, the details of a ‘healthy’ microbiota remain poorly

defined (7). Spatial and chemical heterogeneity of the habitat adds considerable complexity to accurate modeling.

Conceptualizing the Cellular Gut Microbiota

Several heuristics have been applied to cellular GI microbiota diversity with varying success. Enterotypes were first proposed in 2011 (31). Using cross-sectional analysis of whole genome shotgun sequences, investigators were able to bin healthy study participants into one of three community types which cluster based on the prevalence of one of three bacterial genera. Subsequent temporal analysis of other subjects revealed that these community states are not stable through time within an individual and are more likely conglomerations of continuous rather than discreet community compositions (32, 33). Another mode of dissecting the GI microbiome involves the identification of common taxa widely observed in people without overt disease. To this end, a core 'healthy' bacterial community defined at the species level and found within more than 50% of healthy individuals surveyed has been described (34). Perhaps the most enduring taxonomical community model for a healthy GI microbiota is the association of high taxonomic diversity with the dynamic notion of host health (4). Unlike phylogenetic comparisons, GI microbiome functional capacity is better conserved across human populations and exhibits a high level of redundancy within healthy subjects (3, 28, 34).

Viruses of the Gut Microbiota

Viruses constitute the second most abundant component of the human GI microbiota with an estimated density of 10^9 to 10^{10} virus like particles (VLPs) per gram

of stool dry weight or per cubic millimeter of mucosal biopsy (35, 36). There is some conjecture that this estimate is low and that the true concentration of VLPs in the colon is 10^{10} to 10^{12} per gram of dry feces (16). These densities equate to a virus – microbe cell ratio between 1:10 and 1:1. Bacteriophage dominate in these viral communities and represent the greatest species diversity within the habitat (37, 38). Other populations include eukaryotic viruses derived from the host, diet, or eukaryotic members of the microbiota (e.g. fungi), and potentially a small number of archaeal viruses though this is currently unsubstantiated (38, 39).

The bacteriophage community is enriched for double stranded DNA phage and single stranded DNA species (38, 40). RNA phage have been detected in very low numbers and diversity (41). Viruses associated with the newly identified Candidate Phyla Radiation may also be contained within the gut microbiome community but rigorous investigation is necessary to advance this notion.

Characterization of the gut microbiome phage community is ongoing. Direct detection using electron microscopy and genomics detection show a considerable number of head tail phage from the order Caudovirales as well as examples of Microviridae (36, 42, 43). A majority of metagenomic phage DNA sequences from the human gut have not yet been classified taxonomically (21, 37, 40). Early studies reported an inability to classify upwards of 80% of metagenomic sequences. This number is improving but still far from saturation. The phage community also appears to be individual specific and stable through time in the absence of perturbation or disease (38, 44). Several studies have uncovered commonly distributed phage genomes within healthy human populations

and a subset of these have been identified as a healthy core sub-community complimentary to the concept of a healthy core bacterial community (40, 43, 45-48). Alterations of the gut phage community can be readily observed in different disease states and as the result of dietary or chemical perturbation (21, 40, 49, 50). Lepage et al. 2008 identified an increased abundance of phage particles in intestinal biopsies of individuals with ulcerative colitis compared to healthy controls (36). A more recent study found that phage diversity, especially within the order Caudovirales was significantly expanded in patients with irritable bowel disease (21). Disturbance of the microbiome with antibiotics also expands active phage diversity and dietary shifts are capable of producing convergent alterations in GI virome composition (40, 49). As with the cellular component, a healthy viral community has not been completely defined and causal relationships between observed alterations in community composition and host condition remain elusive. The key paradigm here is that viruses are a major component of the ecosystem. Though often viewed through the lens of pathogenisis, their true role appears to exist on a continuum from homeostasis to dysbiosis. One of the outstanding questions facing the field is: what role, if any, do the viruses associated with the gut microbial community play in heath and disease?

Experimental Approaches to Phage in the GI Microbiota

Many platforms have been either developed or adapted to study bacteriophage ecology in humans. As far back as the 1950's, animal models were used to analyze the biogeography of phage throughout the metazoan host (51, 52). These studies showed that phage are common in host tissues and can retain viability when transcytosed and

transported throughout the host. Humans and various animal models including gnotobiotic ‘humanized’ mice have been used to investigate the response of the viral community to antibiotics, other xenobiotics, host genetics, and diet (40, 49, 53-55). Reyes *et al.* used gnotobiotic mice colonized with a 15 member model bacterial consortium to capture phage from human stool and observe phage host interactions over several weeks (56). They clearly observed lytic, kill-the-winner dynamics as well as potential epigenetic modification and lysogenic relationships. Here, kill-the-winner refers to the predator-prey interaction model proposed by Alfred Lodka and Vito Volterra in which increasing host density correlates with increased predation which in turn diminishes host density in cyclic regime (57, 58). Others have observed colonization resistance, or the inability of exogenous species to establish within the gut, and phage genome integration into the bacterial genome (lysogeny) in similar model studies (59, 60). During study periods, typically days to weeks, animal models only offer limited sampling opportunities which are generally restricted to fecal samples. Fecal samples, unlike site-specific samples are a mosaic of interactions which occur throughout gut transit.

Multiple *in-vitro* methods have been developed to examine gut microbiome-phage interactions. *In-vitro* methods include tissue culture systems like the Gut-on-a-Chip (61). Gut-on-a-Chip involves culturing epithelial cell layers in the flow channel of a peristaltic chip. Barr *et al.* successfully used this system to test their Bacteriophage Adherence to Mucus (BAM) immunity theory (62, 63). The authors demonstrated that phage can create a high localized density in mucus do to sub-diffusive motion and potentially protect the metazoan host from bacterial invasion. They also corrected the assumption that certain

phage preferentially adhere to mucus via immunoglobulin like capsid domains. A more recent *in-vitro* study used cell monolayers to investigate the potential mechanism of transcytosis from the external gut to the internal environment showing a preferential migration from apical to basal cell surfaces and potentially explaining the etiology of phage in host tissues (64). To my knowledge, *ex-vivo* systems such as intestinal organoids have not yet been applied to the study of phage in GI microbiota.

Culture based techniques were the first employed to study phage isolated from the gut (65, 66). These studies involved a limited number of cultivable host species and isolated phage which did not reflect the complex diversity of the gut microbiome. Since the advent of genetic engineering, phage research has focused on molecular applications and investigation of phage ecology has been less emphasized. However, isolate culture techniques are frequently employed alongside *in-vivo* studies (49, 53, 67).

In recent years there has been advancement in the use of culture-based assays that support a richer microbial consortium. Community culture approaches have been shown to mimic the distal colon bacterial community (68-71). Depending on donor communities, much of the original taxanomic diversity can be maintained. Depending on analysis this can mean maintenance of several hundred species level groupings (39, 68, 71-73). These platforms have been employed to investigate physiological aspects of the gut microbiota across host disease states (23, 74, 75) (76). They offer the advantage of direct, real time sampling where clinical and animal models limit sampling to fecal material unless invasive procedures are employed. Recently Santiago-Rodriguez *et al*. combined next generation sequencing and continuous culture to investigate the ability

chemostats seeded with human fecal communities to maintain relevant phage populations (39). They found that humanized chemostats can support a diverse viral community and maintain the interpersonal variation observed in the original donors.

Summary and Project Hypothesis

Parasites play a major role in the assembly and function of ecosystems. Within the human GI tract microbial viruses, and bacteriophage in particular, are abundant elements which likely influence the maintenance of homeostasis as well as its degradation. The summation of phage-host interactions at the community level is no doubt a non-linear function which challenges classic molecular techniques of isolation and interrogation. Furthermore, interpretations within a host associated context are complicated by the spatial heterogeneity of the gut habitat, host immune function, genetics and environmental variables.

While a growing body of observational work exists which has principally established correlative models, *in-situ* experimentation on human associated microbiotas is limited. Ethical issues surround unencumbered experimentation on human subjects and the need to satisfy rigorous empirical standards limits inferential power. That said, further observations at the ecosystem level are essential to designing relevant studies for experimental platforms.

In- vitro, -vivo and *ex-vivo* techniques each offer specific advantages at the cost of certain incongruities. Continuous culture systems that support a large consortium of gut microbial members are imperfect but potentially useful tools for deciphering community

level phage-microbe relationships. The following chapters explore this possibility. The specific hypothesis to be tested are the following:

Hypothesis 1:

Continuous culture systems seeded with human fecal microbiota harbor considerable bacteriophage diversity including phage taxonomies commonly found in healthy humans.

Hypothesis 2:

Extracellular elements, especially phage, impact the cellular community structure of the human GI microbiota.

Experimental Approach

To address the first hypothesis, bioinformatic observations based on metagenomic data were used. A database generated from longitudinal fecal virome sampling of two Montana State University donors was used to investigate the occurrence of common viruses within fecal chemostat cultures. Viromes from two continuous culture systems were used for the comparison. Sequence data from fecal chemostat viromes published in Santiago-Rodriguez et al. 2015 were mined and new metagenomes were generated from reactor material provided by collaborators at the Baylor College of Medicine.

The second hypothesis was tested employing a minibioreactor continuous culture system. Viruses were enriched from donor fecal samples and later applied to stable reactor communities. Alterations in bacterial community structure were monitored.

Limitations and Assumptions

Consensus is currently lacking on the appropriate metrics for assessing the role of viruses within microbial communities (77). ‘Omic’ techniques allow community and even ecosystem scaling of molecular analysis with coinciding losses of precision. Some evidence indicates that virus-host relationships can be deciphered through paired viral and cellular metagenomic time-series data (78). Sequence analysis and binning techniques like average nucleotide identity (ANI) show some ability to infer virus-host associations from static datasets (79). Ultimately, however, true associations and dynamic interactions remain unsubstantiated for the majority of microbe-virus pairings. The investigations described here rely on a combination of targeted quantitative Polymerase Chain Reaction (qPCR), 16S ribosomal RNA based amplicon community profiling, and viral shotgun metagenomics. As collected, these data offer the ability to monitor bacterial community structure at approximately sub-genera resolution, individual virotype dynamics, and input viral community genomic composition. Key assumptions include the predominance of bacteria and their viruses both within the host and in cultured communities, and the ability of 16S rRNA V4 community profiling to detect relevant community shifts in response to viral challenge. Inference to specific virus host interactions is not feasible with this particular data and, moreover, is considered highly tenuous via post hoc analysis.

Chemostat limitations

Interactions of specific virus-host pairs can be altered between the *in-vivo* digestive tract and *in-vitro* culture environments. For the purpose of this research, we assume that the fundamental mechanics of many phage-host pairs are strong enough to preserve certain canonical aspects of their relationship from a host associated context.

Isolation of viral communities from stool is a difficult proposition. Within double stranded bacteriophage, size and chemical stability vary widely (80). Adherence to mucus, cells, and a milieu of other organic matter within feces also complicates the situation. Our protocol to collect viable phage fractions from human stool samples attempted to minimize processing steps, dilution, mechanical shearing and size exclusion. A range of eukaryotic viruses, and small molecules including peptides, and vesicles are certainly contained within the phage fractions. Both spores and small cells like those found in the newly described Candidate Phyla Radiation could feasibly be contained in the filtrates, however, the gathered data does not appear to support this. For our purposes we assumed that bacteriophage were the most enriched elements within the phage fractions and the most likely to impact established bacterial bioreactor communities.

CHAPTER TWO

BACTERIOPHAGE COMMUNITIES IN CONTINUOUS CULTURE SYSTEMS

Hypothesis

Continuous culture systems seeded with human fecal microbiota harbor considerable bacteriophage diversity including phage taxonomies commonly found in healthy humans.

Investigations addressing Hypothesis

The suitability of continuous culture systems to support a diverse bacteriophage community from healthy human GI microbiota was conducted. Viral metagenomes from a humanized continuous culture bioreactor array were generated in collaboration with researchers at the Baylor School of Medicine. The resulting viral metagenomes were analyzed along with previously published viromes from independent studies by Santiago-Rodriguez *et al.* and Manrique *et al*. (39, 46). Some attention was also given to the DNA sequencing and analysis strategy employed by Santiago-Rodriguez.

Approach I: Data Mining Santiago-Rodriguez *et al.* 2015

Overview

Previously, Santiago-Rodriguez *et al.* reported on human colonic viral communities maintained in chemostat cultures (39). The investigators found that chemostat cultured communities maintained viral populations of similar density to those

in donor feces (~10^9 particles per milliliter). Interpersonal variation of viral communities was also maintained and chemostats were able to retain a subset of phage species present in the host colon community.

The investigators used Ion Torrent semiconductor sequencing to characterize viral DNA extracted from cesium chloride purified viral particles from five human donors and associated longitudinal chemostat samples across 24 days of operation. At a sequencing depth of only approximately 500,000 reads per sample and average read length of 215 base pairs this effort likely only captured a fraction of the gene content present in each virome sample. In terms of sampling diversity however, the authors utilized a contig spectra clustering method, the Homologous Viral Diversity Index or HVDI, which was satisfactorily saturated when rarefied to as low as 20,000 reads (81). As an extension of these findings, I attempted to identify the presence of any Healthy Gut Phageome (HGP) contigs derived from two well characterized donors at Montana State University and categorized based on their occurrence within other healthy adult fecal viromes (46).

Methods

Chemostat sequences were retrieved from the Metagenomic Rapid Annotations using Subsystems Technology (MG-RAST) server (82). These included virome sequences for five donors from five culture time points as well as the original fecal sample. All five donors were healthy people both, male and female, recruited through the University of Guelph in Ontario Canada. They ranged in age from 7 to 44 years. Three constituted a family group. The researchers controlled for antibiotic use requiring no administration for at least 9 months prior to the time of sampling. These files were

composed of the quality filtered reads used for assembly in the original investigation. Residual barcode and adapter bases were trimmed from reads prior to assembly. Contigs were generated using the IDBA-UD assembler with modified k-mer criteria (minimum: 100, maximum: 300) and a minimum length cutoff of 1000 base pairs (83). Assemblies were composed of either all chemostat time point reads for individual donors or all reads including fecal viromes for each donor. Quality of assembly was noted based on percent of reads utilized, and assembly length profiles.

Raw reads were de-duplicated and trimmed to a minimum length of 200 base pairs and maximum length of 358 base pairs using Geneious v 10.1.3 default parameters resulting in between 1.1 and 1.8 million reads representing 5 time points and fecal input viromes per donor (84).

Processed reads from across each individual donor's reactor time points were combined and mapped against an MSU contig database consisting of 4301 contigs from two individual's fecal viromes sampled twice over the course of a year (46). The two MSU fecal donors were healthy adults age 26 and 55 years residing in Bozeman, Montana. Neither had consumed or applied antibiotics for at least two months before sampling. This database contained the HGP contigs. It is important to note that the two MSU individuals were entirely external to the chemostat experiment. Reactor reads were combined with corresponding donor fecal viral reads and mapped against the same database. Mapping was carried out in Geneious v10.2 with Bowtie2 high sensitivity, end-to-end parameters (84, 85). Results were manually checked for pair-wise identity of at least 90%. Summary results as well as contig classification based on Manrique et al. 2016

are reported (Table 1; Figures 1, and 2). The process was repeated using the recently described Crassphage genome as a positive mapping target (45). As a negative control, reads were also mapped to a database of Yellowstone National Park hot spring viral contigs using the same parameters. These control contigs generally represent genetic material from thermophilic archaeal viruses and should have little homology to the bacteriophage dominated GI associated viromes.

Results

Modified k-mer cross-assembly parameters in IDBA-UD generate fewer contigs of considerably longer length than those generated with IDBA-UD in the original publication (Table 1). These new assemblies were more comparable to the best published assemblies generated with CLC Genomics Workbench. Between the two IDBA-UD strategies, maximum contig length increased on average 168.5% across samples. Additional assembly statistics are displayed in Table 2 and original Santiago-Rodriguez *et al*. statistics are included in APPENDIX E.

Quality trimmed, length filtered, unique reads from the Santiago-Rodriguez *et al.* chemostat samples mapped to between 134 and 304 MSU contigs with a minimum 90% pair-wise identity. From 1.6% to 25.3% of reads in each sample mapped to an MSU contig (Figure 1). Contigs identified by Manrique et al. as ‘core’ (occurring within >50% of healthy adults surveyed), common (occurring within 20-30% of healthy adults surveyed) and low-unique(found in >20% of healthy adults surveyed) were among those identified by read mapping (Figure 2). Surprisingly, independent mapping against crAssphage produced no hits. This is surprising since the crAssphage family is a globally

distributed hallmark of the enteric virome and would be expected to be present in at least some of the original donor samples if not their associated reactors (45). This may be a limitation of the technique employed more than a lack of crAssphage and related viruses. Interpersonal and geographically driven local adaptations of crAss-like phage may result in greater protein homology with less conserved DNA sequence. Recent investigations indicate that detection crAssphage and highly related viruses can be aided by protein cluster homology analysis (86). Negative control hot springs datasets produced positive matches several orders of magnitude lower in abundance than those produced by mapping to HGP contigs. Specifically, 234 of 4,229,034 reads mapped to hot spring control contigs while 213,068 reads mapped to the MSU contig database.

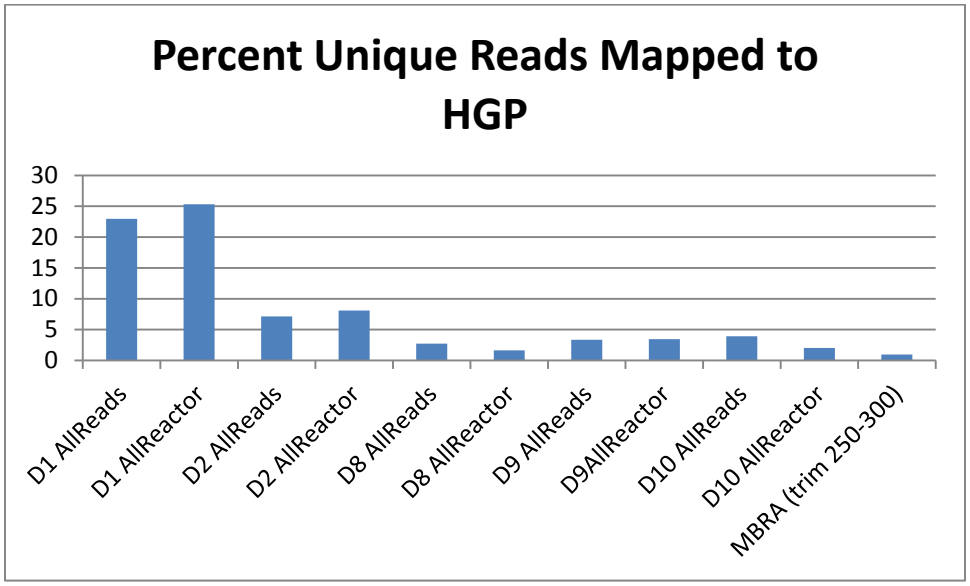

Figure 1: Chemostat reads mapped to HGP. Unique chemostat reads mapping to HGP contigs as percent of total de-duplicated reads per sample.

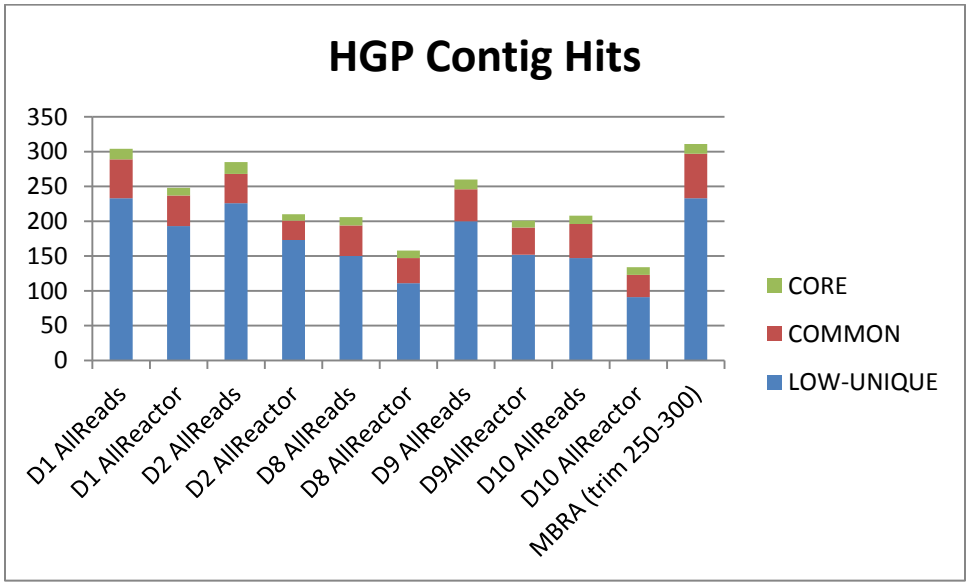

Figure 2: Classification and number of HGP contigs with mapping hits from the various chemostat samples.

	Sample	Contigs CLC	Max Length CLC	Percent in Contigs CLC	Contigs IDBA	Max Length IDBA	Contigs IDBA-UD Modified	Max Length IDBA-UD Modified	Percent in Contigs
Donor 1	Day4	5457	41739	94.5	7883	16701			
	Day8	3280	60551	97.4	5624	20749			
	Day12	3926	53146	95	5199	28300			
	Day16	4064	57967	96.6	6433	18548			
	Day24	3383	63171	96.9	5633	20249			
	Stool	1879	29621	96.1	2830	17331			
	All Reactor						521	60009	42.5
	All Reads						634	70499	45.2
Donor 2	Day4	1946	38124	96.8	3246	18417			
	Day8	4206	25757	96.6	6925	16493			
	Day12	3201	33183	95.5	5912	12411			
	Day16	5698	53789	97	9203	19208			
	Day24	6022	51019	96.2	9030	15057			
	Stool	3739	52910	92.4	4188	27868			
	All Reactor						527	59067	46.2
	All Reads						1125	69476	46.2
Donor 8	Day3	6541	49980	95.3	6723	15285			
	Day6	4547	43340	95.3	3693	18075			
	Day12	5781	49984	95	6096	16130			
	Day18	1114	48541	98.2	3674	21185			
	Day24	1018	48504	98.8	3299	15020			
	Stool	1266	19694	96.6	3576	14004			
	All Reactor						410	76627	41.7
	All Reads						511	76630	44.1
Donor 9	Day3	3391	69536	97	4732	21652			
	Day6	2425	45528	98	4271	34999			
	Day12	3556	94115	97.1	3749	13218			

Table 1: Assembly comparison for Santiago-Rodriguez et al. data. Original statistics adapted from publication for comparison purposes.

Table 1 Continued

Donor 9 continued	Day18	2867	93797	97	3672	13588			
	Day24	2913	93414	96.5	3332	9701			
	Stool	2696	41934	94.7	3090	32413			
	All Reactor						327	71393	44.2
	All Reads						484	93790	44
Donor 10	Day4	4533	51727	96	8707	16925			
	Day8	5148	40455	97.1	6454	17015			
	Day12	5891	48332	96.1	6643	29283			
	Day16	5365	40497	96.2	5674	14373			
	Day24	4808	40172	97.9	7070	17794			
	Stool	3364	50410	97.6	3264	27849			
	All Reactor						1677	51354	34.2
	All Reads						1206	62863	40.8

Assembly	Contigs	n50	Max	Mean	Total Length	n80	Reads	Reads Aligned	Percent
Donor1 AllReactor	521	9421	60009	4598	2395739	2236	2509031	1066341	42.50011
Donor2 AllReactor	527	5184	59067	3703	1951999	1962	2443339	1127260	46.13605
Donor8 AllReactor	410	11386	76627	4431	1816854	2057	2312886	964757	41.71226
Donor9 AllReactor	327	5963	71393	4033	1319115	2017	2380987	1053295	44.23775
Donor10 AllReactor	1677	4337	51354	2499	4191607	1740	2467899	843568	34.18163
Donor1 Allreads	634	10741	70499	4934	3128440	2474	3737968	1688123	45.16152
Donor2 Allreads	1125	6255	69476	4211	4737547	2222	3864218	1785169	46.19742
Donor8 Allreads	511	10168	76630	4630	2366071	2255	3688775	1627434	44.11855
Donor9 Allreads	484	7060	93790	4248	2056057	2202	3487768	1533455	43.96666
Donor10 Allreads	1206	6718	62863	4467	5387235	2468	3805875	1552699	40.79743

Table 2: Additional assembly statistics for modified IDBA-UD assemblies. '-_AllReactor' indicates a cross assembly of all chemostat timepoints from Santiago-Rodriguez et al. dataset. '-_Allreads' indicates cross assemblies including all chemostat samples and associated fecal samples.

Assembly	Percent Increase largest contig
Donor 1	149.1131
Donor 2	149.3039
Donor 8	261.7182
Donor 9	167.9791
Donor 10	114.674
Average	168.5577

Table 3: Longest contig comparison. Increase in length for largest contig generated with modified IDBA-UD assembly compared to that reported for standard IDBA-UD assembly in Santiago-Rodriguez et al.

Discussion

The Santiago-Rodriguez dataset can be challenging from an assembly standpoint. Single end semiconductor sequencing lacks paired end orientation information. GI viral diversity is considerably uneven and can be much richer than the associated cellular community. This increases the need for adequate sequencing depth which, at ~ 500,000 reads/sample, was below published collector's curve saturation levels of 10^7 Illumina HiSeq reads for a human viral metagenome (44). At on third the average length of the Santiago-Rodriguez reads this would translate to a saturation point around 3 million Ion Torrent reads. It is also important to note that the benchmark study required a far greater sequencing depth to generate the contigs used as rarefaction targets as well as a larger minimum length cutoff of 1kb per contig.

Our modified IDBA-UD and cross-assembly techniques both improved the output of the IDBA assembler when compared with the previous Santiago-Rodriguez *et al.* published results. With respect to phylogenetic modes of community analysis, this is a considerable improvement for the IDBA output. In all donors, contig size increased with maximum contig length increasing over 100% in each donor. A similar improvement does not necessarily hold in terms of analyzing total virome coding potential. Santiago-Rodriguez *et al*. were able to include a greater portion of the total sequenced genetic content by applying their Homologous Viral Diversity Index (HVDI) to assemblies lacking a stringent length criteria. This was also due in part to the authors using a relaxed minimum length cutoff for contigs. Newer metagenomic pipelines, including this analysis, commonly set a minimum contig length cutoff of at least 1000 base pairs. A

cutoff in this range minimizes the impact of chimeric artifacts and is long enough to include at least one average sized protein coding sequence per contig. Sufficiently deep sequencing of communities with variable diversity allows for a majority of the sequenced genomic content to be included in longer contigs. However, it appears that the Santiago-Rodriguez study was able to bootstrap lower coverage samples with the HDVI approach. As both the Young lab and the Pride group have noted, improved assembly strategies with contig clustering represent an efficient way to analyze viromes while lowering sequencing effort. This could also prove useful in rendering comparable outputs from otherwise incompatible metagenomic studies.

Read mapping revealed notable overlap between cultured viromes and the HGP. This adds weight to the observations and conclusions of the original authors by illustrating that not only do chemostats retain a portion of the original donor phage community but they are able to support commonly occurring viruses of the community which have been correlated with health. These observations do not imply that specific interactions between phages and hosts are strictly maintained in culture but do provide rationale for further investigation.

MBRA Viral Screening

Overview

Within this thesis, minibioreactor arrays, or MBRAs, refer specifically to a parallelized array of continuous culture vessels adapted for the study of human intestinal microbiota by Dr. Rob Britton's research group at the Baylor College of Medicine (68,

69, 87). After examining multiple *in-vitro* study designs, a collaboration with Dr. Britton and Dr. Auchtung was proposed and established thanks to Dr. Seth Walk (MSU). The resulting collaboration generated a majority of the experimental data reported here.

MBR arrays can maintain diverse microbial communities representative of fecal input communities (68, 69, 87). Based on 16S rRNA amplicon metagenomics, as many as 200 species level taxa can be maintained for periods up to several weeks. These communities remain donor specific during operation, recapitulating the inter- and intra-individual dynamics found in human communities through time. Certain clinical aspects of the microbiota are also retained such as colonization resistance to *Clostridium difficile* establishment which can be disrupted by antibiotic application. Alongside the results from Santiago-Rodriguez *et al*., this system appeared to be a good candidate for *in-vitro* investigation of GI phage ecology. As an initial step, we screened a set of reactor samples for the presence of phage and overlap with the HGP as we had done with the Santiago-Rodriguez data discussed earlier in the chapter.

MBRA Design

MBRA design is based on previous publications using continuous culture to investigate polymicrobial communities (68, 69, 87). An advantage of this platform is the capacity for technical replication while limiting consumption of materials and space. Individual bioreactors blur the line between biological and technical replication as each separate reactor will invariably diverge from the initial community, however these approximations still represent an analytical yardstick in ecological investigations.

MBRA vessels operate at a volume of 15 milliliters, a fraction of that used in the Santiago-Rodriguez study. Flow rate is maintained at 1.875 milliliters per hour corresponding to a wash out time of approximately eight hours. Media composition also varied between the two reactor systems in previous experiments however this can be easily manipulated in future work.

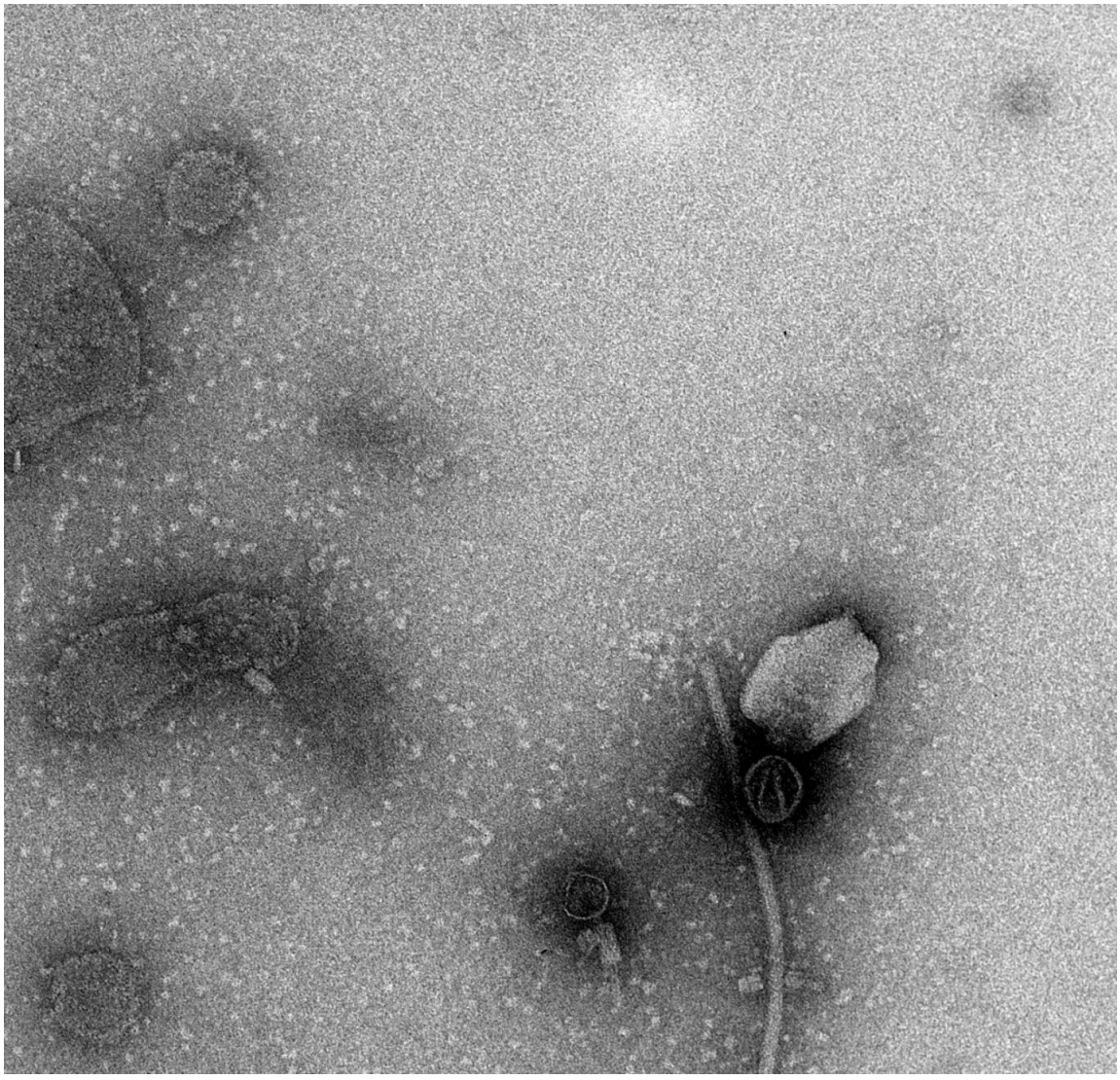

Figure 3: TEM of MBRA Concentrated Supernatants. This show a concentrated supernatant magnified 20,000 times. Many small virus like particles are visible including at least one head tail morphology.

Methods

To characterize the viral community maintained in the MBRA system, endpoint material from a study carried out by colleagues at Baylor College of Medicine was analyzed. Reactor communities were allowed to stabilize for eight days at which point replicate groups of three were treated with one of three antibiotics or a sham control

twice daily for three days then allowed to recover for two additional days before being challenged with *C. difficile* spores. *Clostridium difficile* colonization was monitored for an additional three days at which point whole reactor contents were harvested, cells removed by centrifugation and both cell pellets and supernatant were flash frozen at -80℃. These materials were shipped frozen to MSU for analysis.

Upon receipt, supernatants were thawed on ice and filtered through 0.8um in-line syringe filters. Virus particles present in the filtrates were concentrated and pelleted by ultracentrifugation at 35,000 x g in a Beckman Colter fixed angle MLA 80 rotor for 2.5 hours at 10℃. Supernatant was removed and pellets were covered with 100ul chilled SM buffer (100mM NaCl, 8mM $MgSO_4 \cdot 7H_2O$, Tris-Cl 50mM (1M, pH 7.5)) and allowed to re-suspend overnight at 4°C. Concentrates were imaged at 20,000 times magnification using 2% uranyl acetate negative staining on a Leo912AB transmission electron microscope (TEM) to confirm the presence of virus like particles. Samples were prepared for shotgun metagenomic analysis by pooling supernatants from replicate reactors to a volume of approximately 6ml before pelleting and re-suspension of viral particles as described for imaging. Viral DNA was isolated using a PureLink Viral RNA/DNA isolation kit with an extended 2 hour proteinase K digestion. DNA concentration was quantified with a Qubit fluorimeter high sensitivity DNA assay, integrity was visualized using gel electrophoresis, and 10ul of each sample was aliquoted for sequencing. Metagenomic sequencing was performed by the University of Illinois Sequencing Center (Champaign-Urbana) using an Illumina MiSeq platform, generating

300bp paired-end reads with V3 chemistry to a total of 25 million reads across six samples (4 MBRA and 2 fecal samples[1]).

Sequences were quality trimmed then filtered for lengths greater than 250bp. *De novo* assembly was performed by modifying IDBA-UD assembler to maximum k-mer lengths of 300bp and augmented read lengths of the same. Additionally, a cross-assembly approach was employed with the aim of capturing low abundance contigs. A cutoff of 1000 bp was applied to all contigs as has been adopted in other metagenomic virome studies. Raw quality trimmed reads from each sample were then mapped to the co-assembly contig database. Open reading frames were predicted and extracted from individual contigs using WebMGA (88). These were queried against the Phage Orthologous Group (POG) database via psiBLAST to determine the existence of phage specific proteins (e-value threshold < 10^{-5}) (89, 90).

All high quality reads between 250 and 300bp's were mapped against the HGP database using Bowtie2 with high stringency, and end-to-end parameters (85). The same one-hit mapping criteria described above was applied. Raw reads were queried against the Silva 16S rRNA database with BLASTn using an e-value cutoff of 1 x 10^{-10} (91, 92) to assess bacterial contamination.

Results

Virus like particles were directly detected in TEM images including examples of head-tail morphologies (Figure 3). Cellular contamination was assessed using the BLAST results of total raw reads against the SILVA 16S rRNA database and returned

[1]These two fecal viromes correspond to the input material used in the phage cross experiment which is described in chapter 3.

approximately 16,000 hits from a total of 8 million paired reads, indicating a cellular contamination level < 0.01%. A BLAST of all ORFs against the POG database yielded 5,529 positive hits from 2072 unique contigs from a pool of 7174 contigs from all bioreactors indicating that these contigs are likely representations of phage genomes. This is not to say that the remaining contigs do not represent phage genomic DNA but it does lend evidence to the classification of the 2072 positive contigs. Using the subset of POGs capable of taxonomical inference, a total of 108 matches were confirmed compared with 115 from established human GI phageomes (unpublished data) giving confidence in the phage contribution to these metagenomic datasets (Figure 4). As expected, among contigs with hits to taxonomically validated POGs, a general prevalence of *Caudovirales* was observed in addition to a paucity of *Microviridae* which is recapitulated in POG analysis of human GI phageomes. Mapping of quality filtered reads against the MSU contig database affirmed the existence of common and core HGP contigs in MBRA samples (Figures 1 and 2). Nearly 40% of the 23 core phageome contigs were detectable within the MBRA sequencing data along 35 of 132 common contigs and a surprising number of less conserved sequences. This degree of representation of MSU database contigs mirrors what was observed in the Santiago-Rodriguez *et al.* dataset and provides additional evidence that chemostat culture systems can host relevant phage taxonomies from the human GI environment.

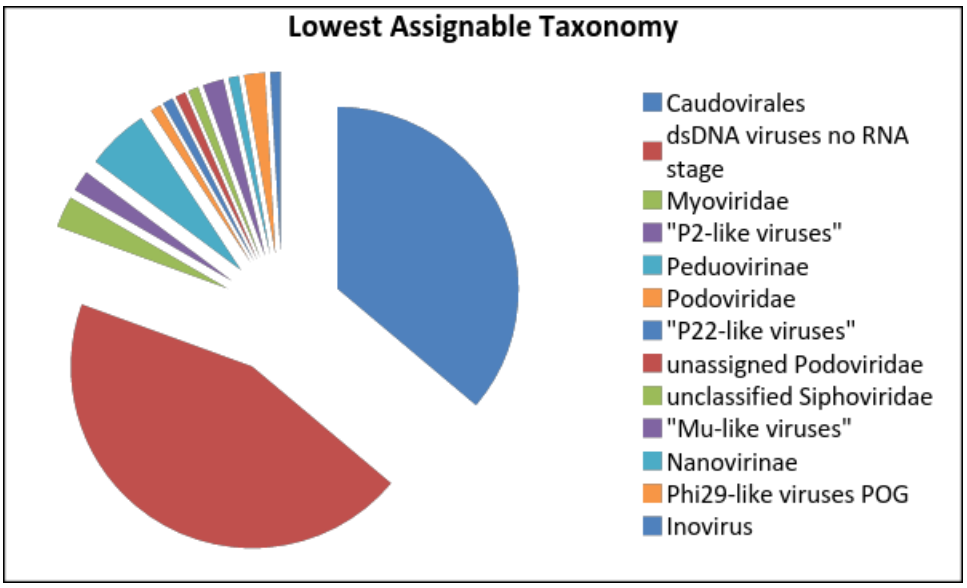

Figure 4: MBRA viral metagenome taxonomies. Proportions of positive POG taxonomy hits to MBRA contig ORFs.

Discussion

The results of these experiments support the hypothesis that the MBRA system can maintain phage communities which are taxonomically similar to those detected in healthy human gut microbiomes. The virome DNA preparation protocol successfully enriched the bioreactor supernatants for bacteriophage while sufficiently depleting bacterial contamination. This method is partially based on unpublished investigations by Pilar Manrique (Young lab, MSU) and allows for comprehensive screening of the viral community without the need for extensive processing steps.

Based on metagenomic analysis, the MBRA system supports a rich phage population. The overlap with MSU donor contigs and the HGP is encouraging. This is especially the case given that the two MSU donors were, to our knowledge, completely independent to the MBRA samples analyzed. Although, it is worth noting that the

sequencing samples were prepared at MSU and there is an improbable but real chance of crossover within the lab.

Technically speaking, MBRA samples were relatively easy to test. Viral fractions were free from excessive particulate and soluble organic material. Viral DNA was able to be extracted with relative ease and produced a non-degraded product for sequencing. The samples were also free of PCR inhibitors which can complicate PCR detection strategies in fecal samples.

Conclusions

The two continuous culture systems analyzed support diverse phage communities. They retain interpersonal variation as seen in human subjects and retain a number of virotypes commonly found in healthy individuals. Despite media and volume differences, both systems captured comparable portions of the HGP. Controlling for substrate effects, MBRA's are likely more advantageous as they require a fraction of the space and consumable material of the larger chemostat systems used in Santiago-Rodriguez *et al.* Based on this investigation and the published works cited, continuous culture systems will likely be useful in disentangling the complex interactions of bacteria and their viruses experimentally.

CHAPTER THREE

IMPACT OF PHAGE ON BACTERIAL COMMUNITIES

Hypothesis

Extracellular elements, especially phage, impact the cellular community structure of the human GI microbiota.

Introduction

Little is known about the role of viruses within the human GI microbiota. Viruses are commonly equated with disease, however, we know they are prevalent in homeostatic microbiomes (37, 38, 46). The overall composition of the bacteriophage community, like the cellular community, is known to vary with host health, diet, and disease state (Figure 5). What remains unknown in this paradigm is the extent to which bacteriophage shape, buffer, or simply respond to changes in the host associated ecosystem. *In-situ* experimentation is limited given the ethical implications, and technical difficulties of applying un-characterized biological elements to patients. In instances where fecal transplantation has been applied, it is extremely difficult to ascertain the specific role of viruses in altering the microbiome separate from the myriad factors at play within the ecosystem. Animal models offer advantages in terms of experimental control and reproducibility but do not mitigate the complex effects of the host or sampling difficulty. Nonetheless, preliminary studies in both humans and animal models have indicated an effect of viruses (and possibility other non-cellular microbiota components), in both

community physiology and host health (49, 53, 93). In an attempt to reduce these complexities and to more directly interrogate the virus-microbe component, we have employed a tractable complex continuous culture approach.

Here we assess the impact of phage on cultured fecal communities by applying autologous (self) and non-autologous (non-self) derived donor phage assemblies. Fecal communities from two healthy human donors were collected and propagated in continuous culture minibioreactor arrays (MBRAs). Reactors from each donor were challenged with viral filtrates obtained from either of the initial stool samples and the bacterial community response was monitored.

Methods

Fecal Inoculum

Fecal samples from two unrelated human donors were collected. Five gram portions were weighed and immediately transferred to an anaerobic chamber where each was homogenized at 25% w/v in SM buffer (100mM NaCl, 8mM $MgSO_4 \cdot 7H_2O$, Tris-Cl 50mM (1M, pH 7.5)). Anaerobic glycerol was added to 15% v/v and slurries were sealed in anaerobic septum vials before freezing at -80°C.

Viral Fraction

From each of the same two donor samples described above, five gram portions were placed in WhirlPAK bags, suspended in 35ml chilled SM buffer and homogenized. The slurries were then transferred to conical tubes and centrifuged at 5,000 x g for 45 minutes at 4^{o} C. Supernatants were carefully collected to avoid disruption of the debris

pellet and filtered through 2um in-line syringe filters. Filtrates were again centrifuged at 5,000 x g for 45 minutes at 4°C. These supernatants were collected and filtered through 0.8um in-line syringe filters. Filtrates were stored at 4°C until application to MBRAs. Viral DNA was extracted and sequenced using 400ul of these filtrates and the process described in Chapter 2 and APPENDIX C.

MBRA Configuration, Phage Spike, and Sampling

The experimental design is illustrated in Figure 6. Bioreactors were prepared as described previously with 6 vessels per donor (3 autologous VLP treated and 3 non-autologous VLP treated) for 12 total reactors (68). Before initiating reactors, 24 ml anaerobic frozen fecal slurry aliquots were transferred to the anaerobic glove box from a -80 C freezer and allowed to thaw. Approximately four milliliters of slurry was inoculated into each reactor. Microbial communities were allowed to establish and equilibrate as batch cultures for 16 hours before flow was initiated at 1.875ml/h and maintained for the duration of the experiment. Three milliliter sample volumes were collected regularly throughout the experiment (time of sampling indicated in Figure 6) and samples were processed according to the following steps. Each sample was divided into two equal portions which were low speed centrifuged to pellet cells. Supernatants were transferred to separate tubes before both cell pellets and supernatants were flash frozen and stored at -80 C prior to further processing. All supernatants and one corresponding cell pellet for each time point were shipped on dry ice to Montana State University. The second cell pellet was retained at Baylor School of Medicine for 16S rRNA bacterial community

sequencing. Anaerobic 15% glycerol stocks were generated for each reactor 6 days after community inoculation and at the end of the experiment.

Seven days after reactor communities were inoculated, VLP preparations were transferred to an anaerobic glove box to reduce overnight. The following day, a 3ml sample was extracted from each reactor. Two hours later a 1ml aliquot of either autologous or non-autologous VLP preparation was spiked into each reactor. Samples were collected over a time course from 2 to 24 hours post VLP addition. This expanding sampling regime was repeated on the second day post VLP addition. Three days (72 hrs) after the VLP addition, all reactor contents were collected (~15ml), two 3ml 15% glycerol stocks were generated for each sample. The remaining culture was split into a 3ml and 6ml portion, cells were pelleted, supernatants separated and both components flash frozen before storing at -80 C.

Individual Phage Dynamics

Individual time point supernatants were filtered through 0.8um syringe filters. DNA extraction was performed as above on 400ul portions. Cell pellets were resuspended in 100ul of corresponding supernatant and DNA was extracted using a MoBio PowerSoil kit. Three primer sets targeting different HGP contigs (Manrique unpublished, APPENDIX D) were assayed against the various MBRA sample preparations and time points using conventional endpoint PCR to control for off target amplification as well as qPCR. The targeted viral contigs were designated virus 25, virus 81, and virus 375.

Target Phage Genomic Comparisons

The two donors' viromes have been previously screened by others at four time points over a period of four years, with the fourth time point representing the MBRA VLP cross input community. To better understand the genetics of the three individual phage targets the most current variations were compared to the historical background for these particular genomes. Reads from the fourth timepoint were mapped to the individual contigs, then mapped reads were used to identify variations from the template sequence.

The relative abundance of target genomes in the bioreactor inoculums and phage spike community was determined. Within each donor's viral metagenome, all reads were mapped against assembled contigs using Bowtie2 with high sensitivity, end-to-end parameters. Relative abundance was then calculated based on the length of contigs and total reads mapped (APPENDIX D).

16S Ribosomal RNA Amplicon Sequencing

This process was also performed at Baylor. Two pre-spike time points and five post spike time points were selected for 16S rRNA community analysis (Table 4, Figure 6). Total DNA was extracted from cell pellets as described previously (68). Bead beating was implemented to minimize differential extraction of gram positive and gram negative bacteria. Amplification and paired end sequencing of the variable region 4 (V4) of the 16S rRNA gene was carried out using the F515/R806 primer set and Illumina MiSeq sequencing protocol.

16S Ribosomal RNA Community Profiles and Statistical Analysis

For the purpose of this project, ordination and multivariate statistics were utilized to determine the basic structural relationships of the sequenced bacterial communities. Forward and reverse amplicon sequences were processed using mothur v1.39.5 with a variation of the MiSeq SOP pipeline (APPENDIX A) (94). Statistical analysis was conducted using R Studio version 1.1.383 using Vegan and Phyloseq packages (95-97).

Principle Component Analysis (PCoA) was applied to samples and the significance of groupings as well as variability were assessed with a combination of Permutation Multivariate Analysis of Variance (PERMANOVA) and beta dispersion. Post hoc analysis is not routine for the adonis function so adjusted pairwise comparisons were generated. Regression was performed against PCoA coordinates and the time series variable.

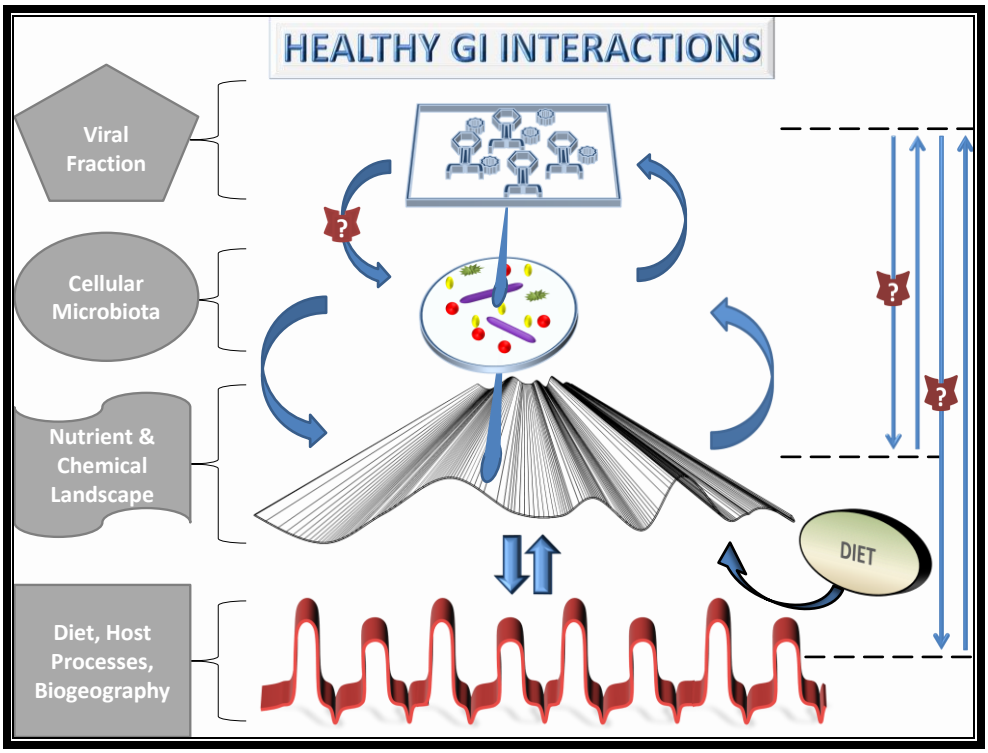

Figure 5: Healthy GI interactions. Conceptual diagram of trophic interactions in the healthy gut ecosystem. Known and proposed interactions are shown with arrows. These can represent either directly and partially-coupled relationships. Question marks indicate ambiguous or unproven interactions

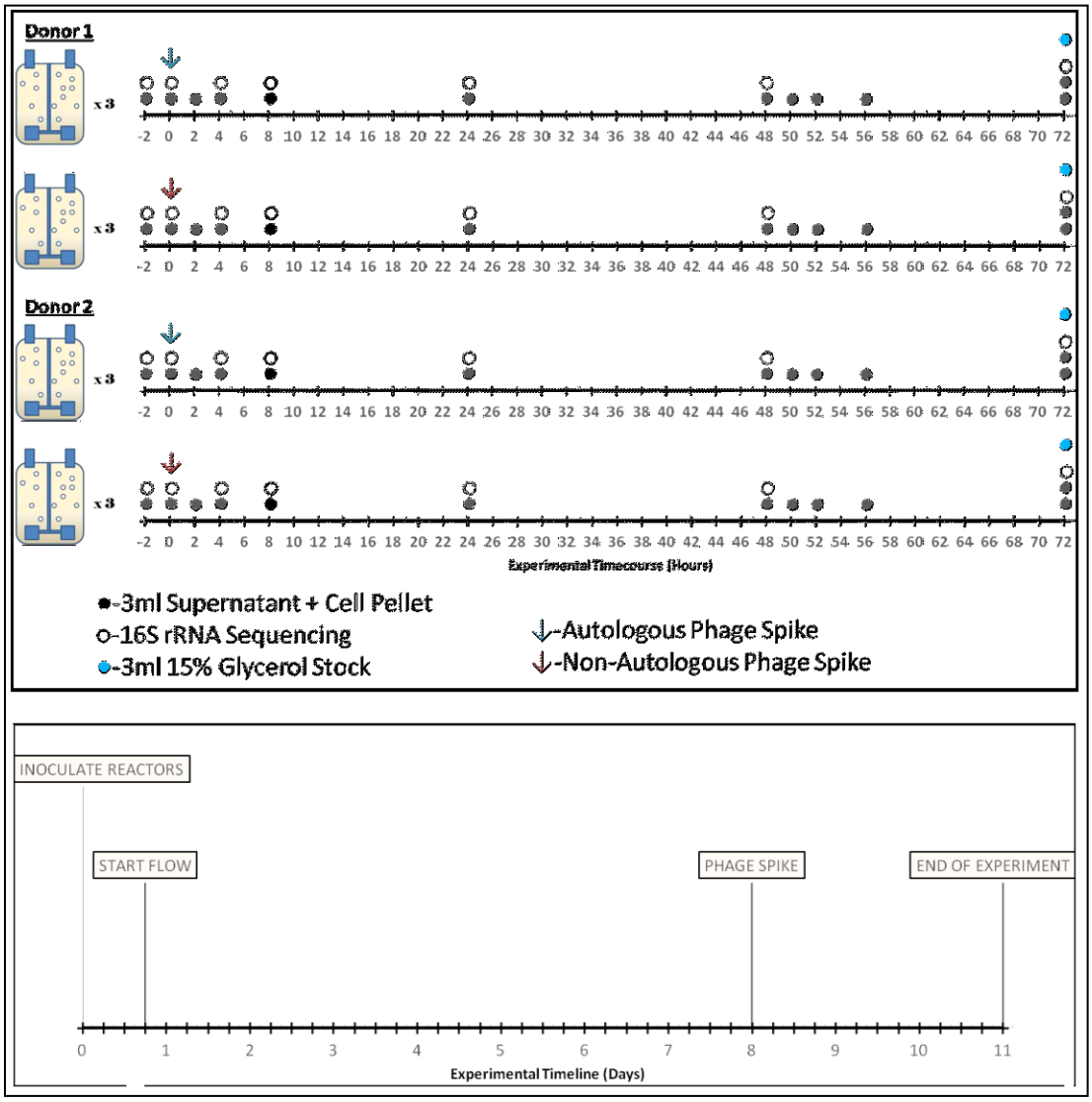

Figure 6: Experimental design. Top – sampling time course immediately before and following phage spike on day 8. Bottom – reactor time course from initial seeding with fecal slurry to end of experiment.

Results

Bioreactor Communities

To identify species level groupings from 16S rRNA amplicons a 97% pairwise identity cutoff was used within the mothur opticlust algorithm (98). This process generated 3455 OTUs across all samples. To collapse potentially spurious diversity, an abundance threshold was applied to each OTU designating the minimum number of sequences required for further analysis. Table 2 illustrates the degree to which various cutoffs collapsed the data. Ultimately a minimum cutoff of 10 sequences was used resulting in 653 species level OTU's. Alpha diversity for reactors is plotted in figure 7.

The bacterial taxanomic composition of bioreactor samples was determined by classifying OTUs against the SILVA 16S rRNA database (91). Community profiles were heavily skewed towards members of the phylum Firmicutes with Bacteroidetes appearing significantly within one reactor from each donor. Lower levels of Proteobacteria, Actinobacteria and Synergistetes were detected throughout the dataset (Figure 8). Pairwise diversity comparisons are illustrated with Bray-Curtis dissimilarity and weighted Unifrac indexes (Figures 9, 10). Reactors clearly align by donor identity. The more diverse donor 1 reactor can be identified as a potential outlier in both heatmaps.

NSEQS (Min seqs/OTU)	**OTU's**
0	3455
10*	653
15	513
25	387
50	254
100	184

Table 4: Total OTU's resulting from various minimum sequence cutoffs. (*) indicates the cutoff used for dowstream analysis.

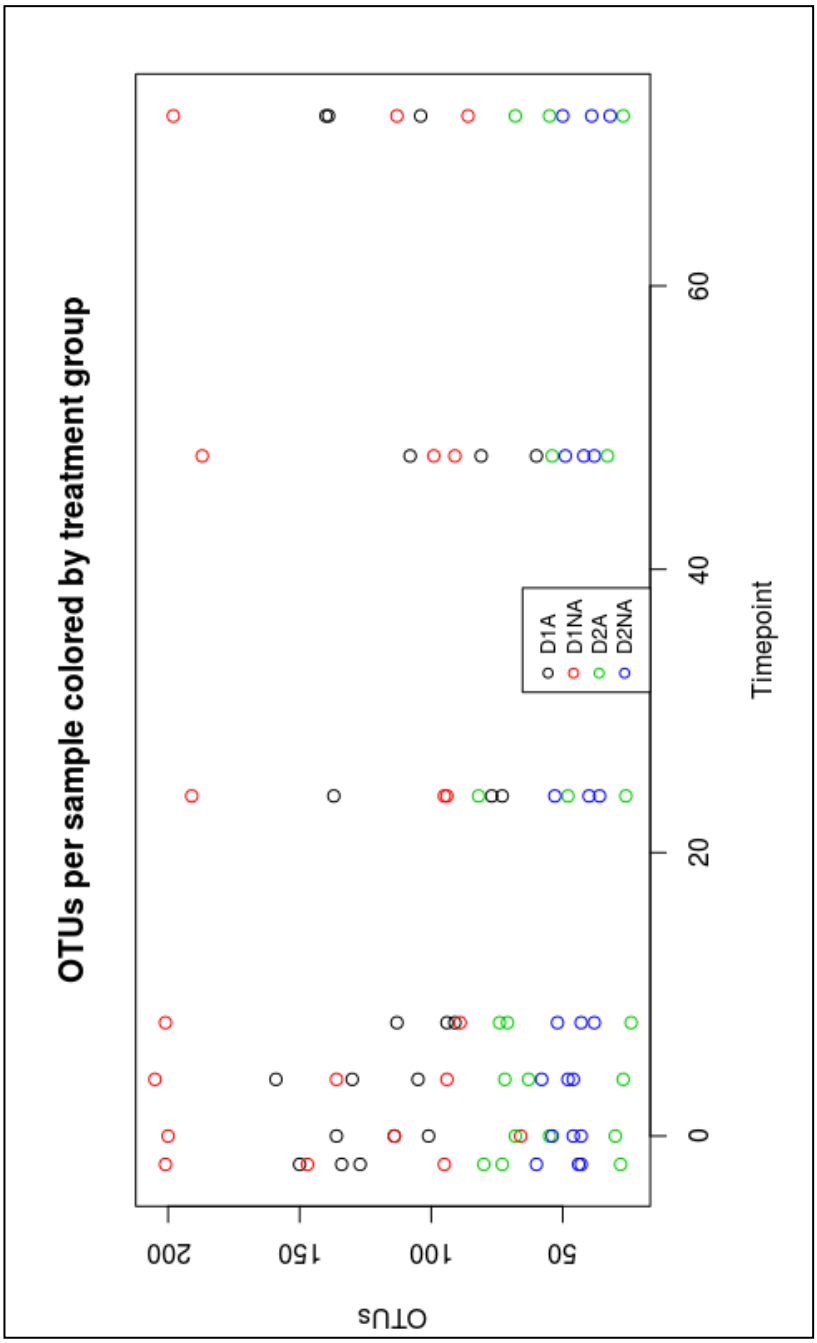

Figure 7: MBRA alpha diversity. OTU's identified throughout 16S rRNA sampling time course. Colors indicate reactor treatment group.D1A – Donor 1 Autologous phage treated, D1NA- Donor 2 Non-Autologous phage treated, D2A- Donor 2 Autologous phage treated, D1NA- Donor 2 Non-Autologous phage treated. Time points reflect hours before and after the phage application.

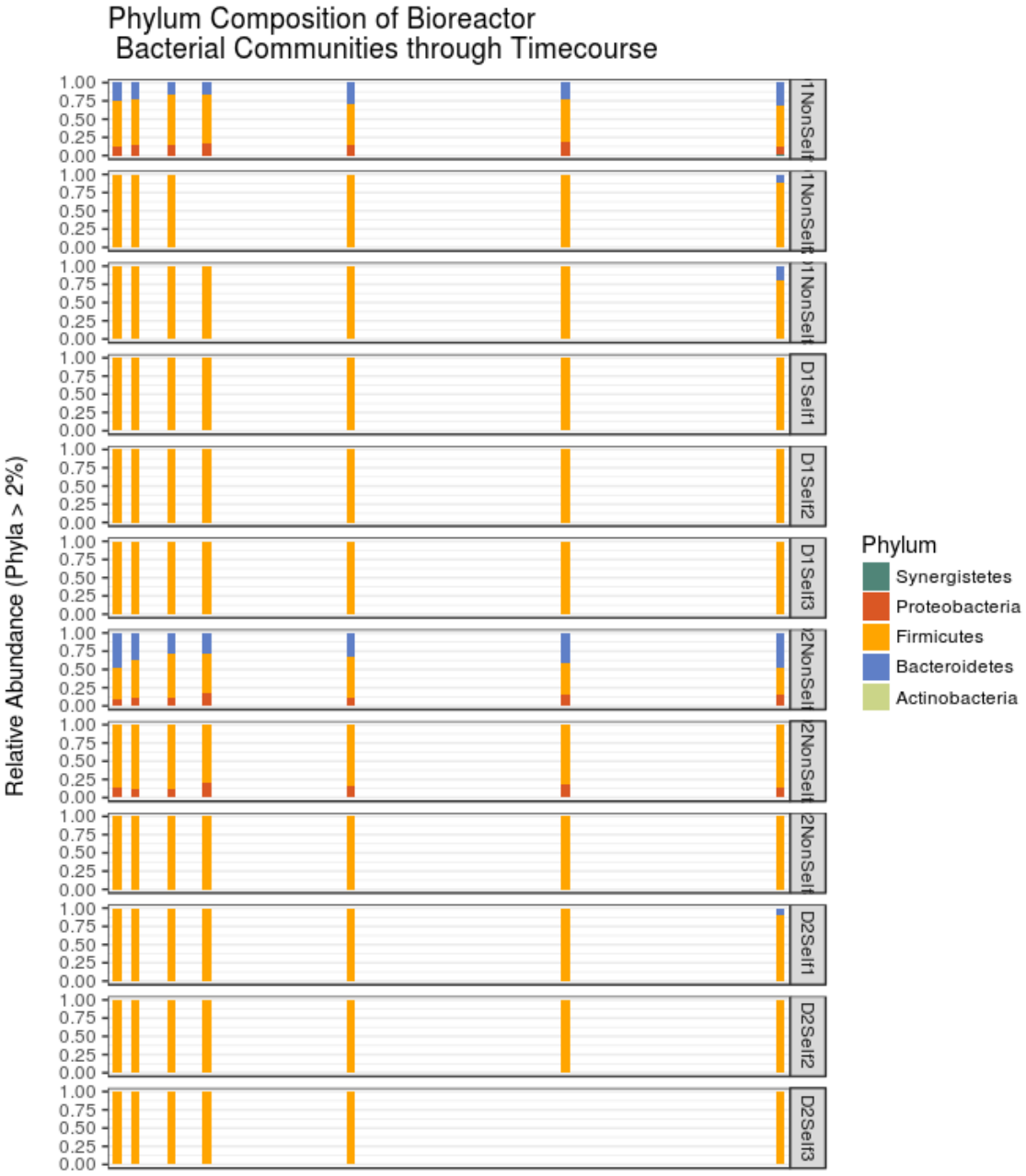

Figure 8: MBRA Phylum diversity. Relative abundance of bacterial phyla within reactors across time course. Only phyla accounting for greater than 2% relative abundance by reads for each time point are included. The time series reflects time before and after phage spike with time points spread on a relative scale.

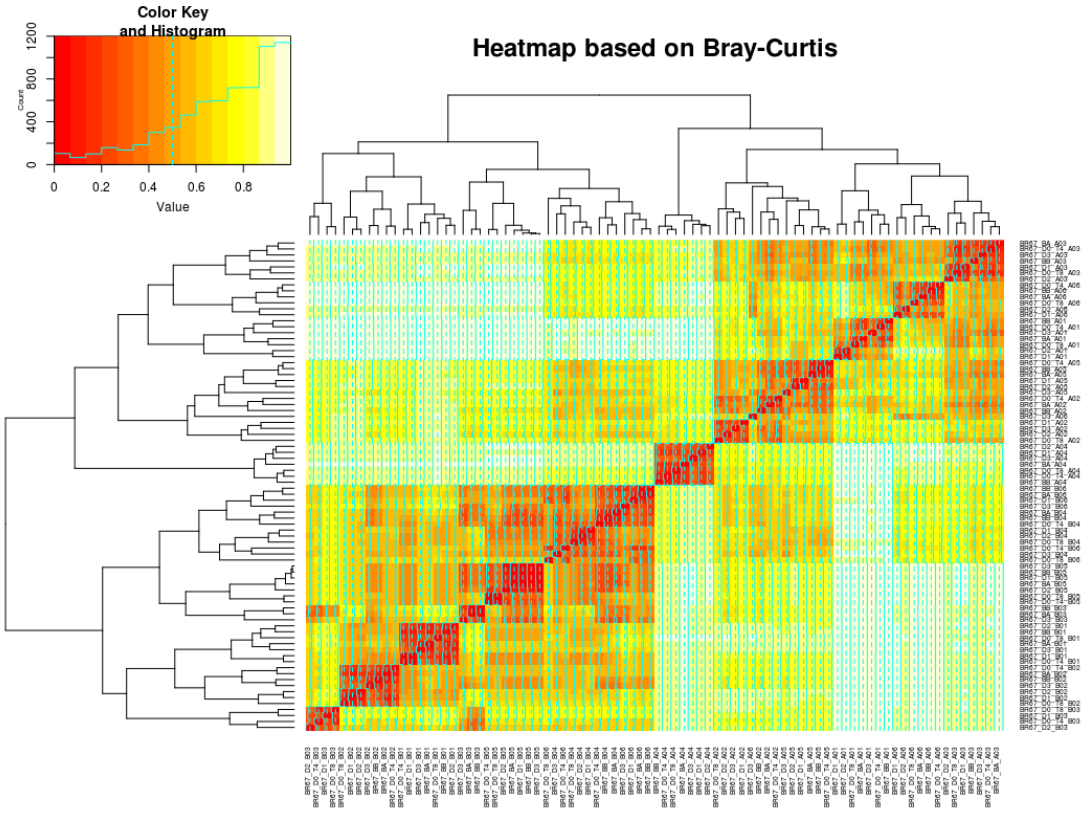

Figure 9: MBRA Bray-Curtis dissimilarity heatmap. Heatmap of pairwise community diversity comparisons for all sequenced samples. Diversity is based on Bray-Curtis dissimilarity index. The similarity based on color is represented in the figure key along with a histogram of occurrences of each similarity level.

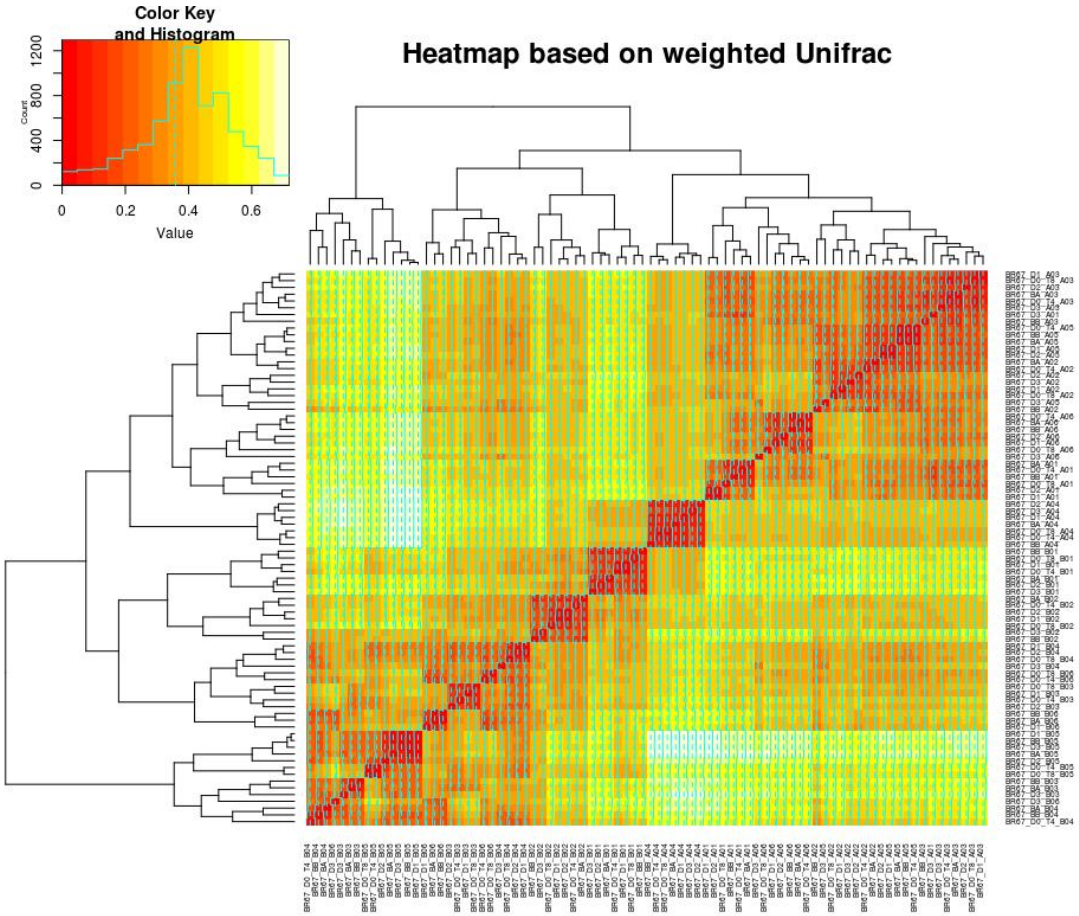

Figure 30: MBRA weighted Unifrac distance heatmap. Heatmap of pairwise community diversity comparisons for all sequenced samples. Diversity is based on the weighted Unifrac index. The similarity based on color is represented in the figure key along with a histogram of occurrences of each similarity.

Viral Dynamics

Three MGEs identified by Manrique et al. 2016 were tracked through the experimental time course using qPCR (Figure 11). These metagenomic contigs represent two complete viral genomes and a mobile plasmid, all found in stool samples of a majority of analyzed human subjects. Virus 81 was known to reside within both MSU donor subjects and was identified in VLP input metagenomes gathered from the fecal

material used to seed the MBRAs. In contrast, virus 25 was present in only one donor at the time of sampling in February of 2017 while plasmid 375 was present in the other donor.

Virus 81 washed out over the seven day establishment period in all reactors. After the establishment period, virus 81 was reintroduced into the MBRAs at the at the phage spike point (time point '0'). The phage spike introduced 10^7 genomes/ml, roughly 10^8 total genomes per reactor, for Donor 2 VLP treated reactors and 10^6 genomes/ml, or 10^7 total genomes per reactor for Donor 1 VLP treated reactors. Donor 2 derived virus 81 persisted in both donor communities for the 3 days following phage spike, whereas Donor 1 derived virus 81 washed out in all communities by 48 hours post phage spike. Compared to the initial titer taken just before reactor flow was started, the Donor 2 VLP spike titers represent one to two orders of magnitude higher concentration of virus 81, indicating that the virus associated with virus 81 was replicating in the MBRA.

Virus 25 was present in Donor 1 stool used to initially establish the MBRAs and the associated VLP extract used to subsequently spike the MBRA. At the time that reactor flow was started it was present at approximately 10^6 genomes/ml in all six Donor 1 reactors, or 10^7 genomes per reactor and was not detectable in Donor 2 reactors. Just prior to phage spike, virus 25 virus titers remained constant between 10^4 and 10^5 genomes/ml across all Donor 1 reactors and remained undetectable in Donor 2 MBRA communities. The phage spike introduced 10^7 particles/ml to all reactors receiving Donor 1 VLP extract for a total of 10^8 virus 25 genomes per reactor. Virus 25 remained present in all Donor 1 reactors detectable between 10^4 and 10^5 genomes/ml for the three days

following the phage spike. In the three Donor 1 non-autologous treated reactors it dropped below detection on occasion but was present in all final samples. The virus associated with virus 25 appeared to successfully establish in the Donor 2 non-autologous treated reactors with a concentration of 10^5 particles in all three at the end of the experiment.

Plasmid 375 was not present in any Donor 1 MBRA reactors prior to phage challenge where it was added at a rate between 10^5 and 10^6 copies/ml before dropping below detection. It reappeared in one of these reactors at extremely low concentration (10^4 copies/ml) in the last four time points. Detection was sporadic across Donor 2 reactors with one non-autologous treated reactor registering high concentrations of virus 375 both before and after the phage spike.

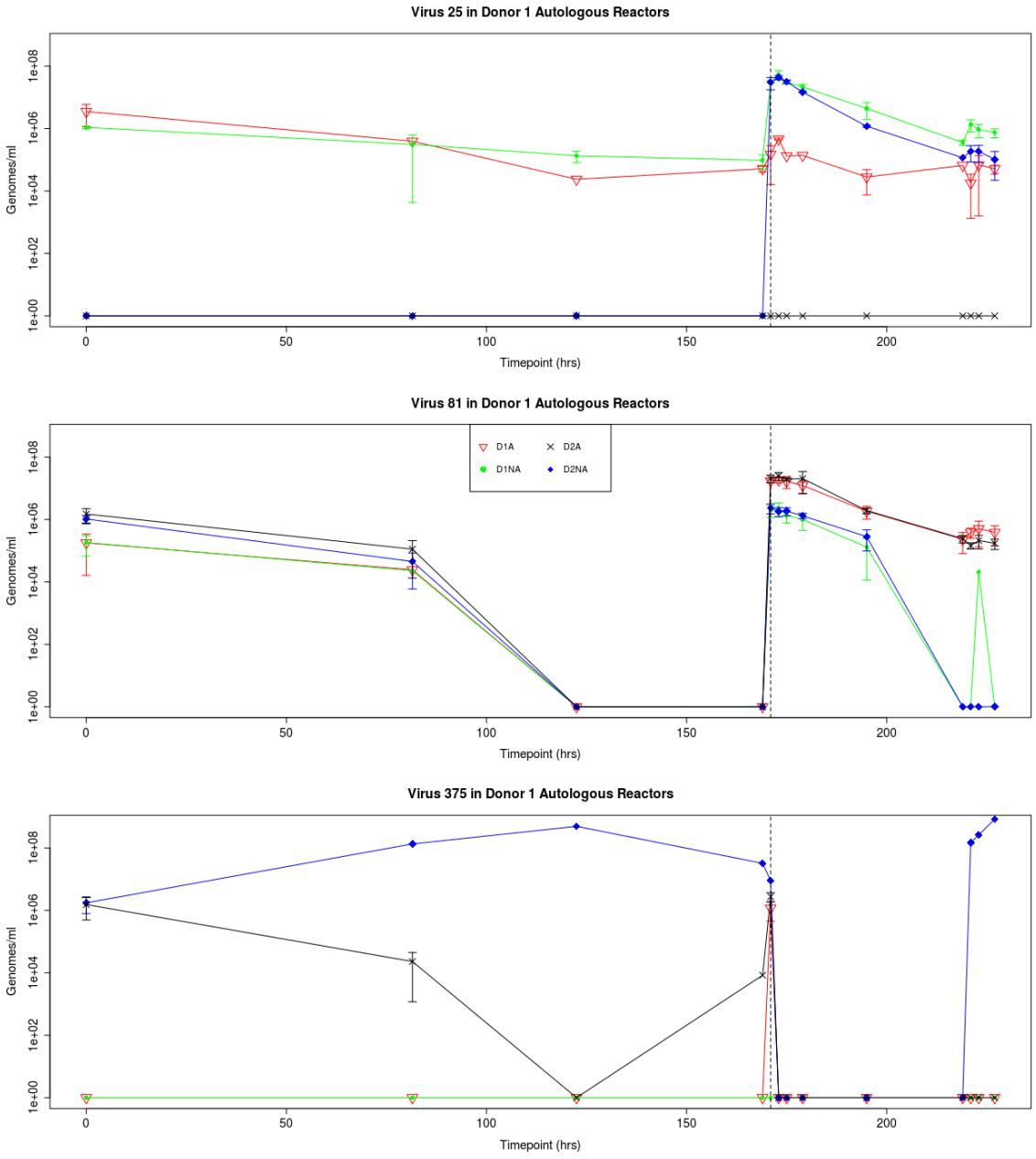

Figure 11: Targeted viral dynamics. Viral titres throughout experiment averaged across replicate reactors recorded in genomes per milliliter. (Green = Not Detected, Red = Highest measured titre). The timecourse begins at the start of reactor flow. Vertical dashed line inticates the time of phage spike. Panel 1 shows results for virus 25 putative phage genome. Panel 2 shows results for virus 81 putative phage genome. Panel 3 shows results for virus 375. Error bars indicate standard deviation across replicate reactors. For points without error bars, only one reactor would have had a positive detection. Various washout and establishment dynamics are apparent.

Bacterial Dynamics

Bacterial communities were first sampled 8 days after reactors were seeded and 7 days after flow was initiated. Individual reactors exhibited divergent structure within donor groupings, however reactors were statistically differentiable based on the original donor (Figure 12) (APPENDIX B). This variation is captured nicely in the primary ordination axis. PERMANOVA results indicate a significant difference in reactor communities based on the associated donor, however the beta dispersion results indicate that this is confounded by differences in variance between the two groups (APPENDIX B) (99). Visually this can be seen in the second ordination axis of the Bray-Curtis based PCoA (Figure 12). Here, one of the Donor 1 reactors is a clear outlier contributing a considerable portion of the total variation in the dataset. Although tempting, removal of the outlying cluster is not experimentally justified as the total replicate number is low and the cluster likely represents expected heterogeneity from a complex environmental sample. The evidence indicating a difference in reactor community composition based on associated donor appears more significant than the confounding affect of heteroscedasticity. For comparative purposes, a separate ordination was generated which omitted the ‘outlying’ reactor (Figure 13).

Donor 1 communities grouped by reactor generate a significant PERMANOVA result when examined at the donor level (APPENDIX B) ($p < 0.001$). With the exception of unweighted Unifrac, the equal variance assumption is met for all other diversity metrics tested (B-disp $p < 0.001$ unweighted Unifrac). The same holds true for Donor 2 samples (PERMANOVA $p < 0.001$).

PERMANOVA analysis of higher order groupings does not contain information about the significance of individual reactor differences, instead only offering the interpretation that at least one of the reactors is significantly different from the others. To better understand inter-reactor variation, pairwise PERMANOVA was conducted for each pairwise combination of reactors. Corrected adonis *p*-values indicate that all but one pairwise combination was significantly different (APPENDIX B). Only two corrected beta dispersion p-values indicated a confounding effect of unequal dispersion ($p < 0.05$). One set, Donor 2 non-autologous phage reactor 3 and autologous phage reactor 2, represented the only case of a difference in variation between the two treatment groups ($p = 0.024$). Donor 2 non-autologous reactors 2 and 3 made up the second significant comparison ($p = 0.042$).

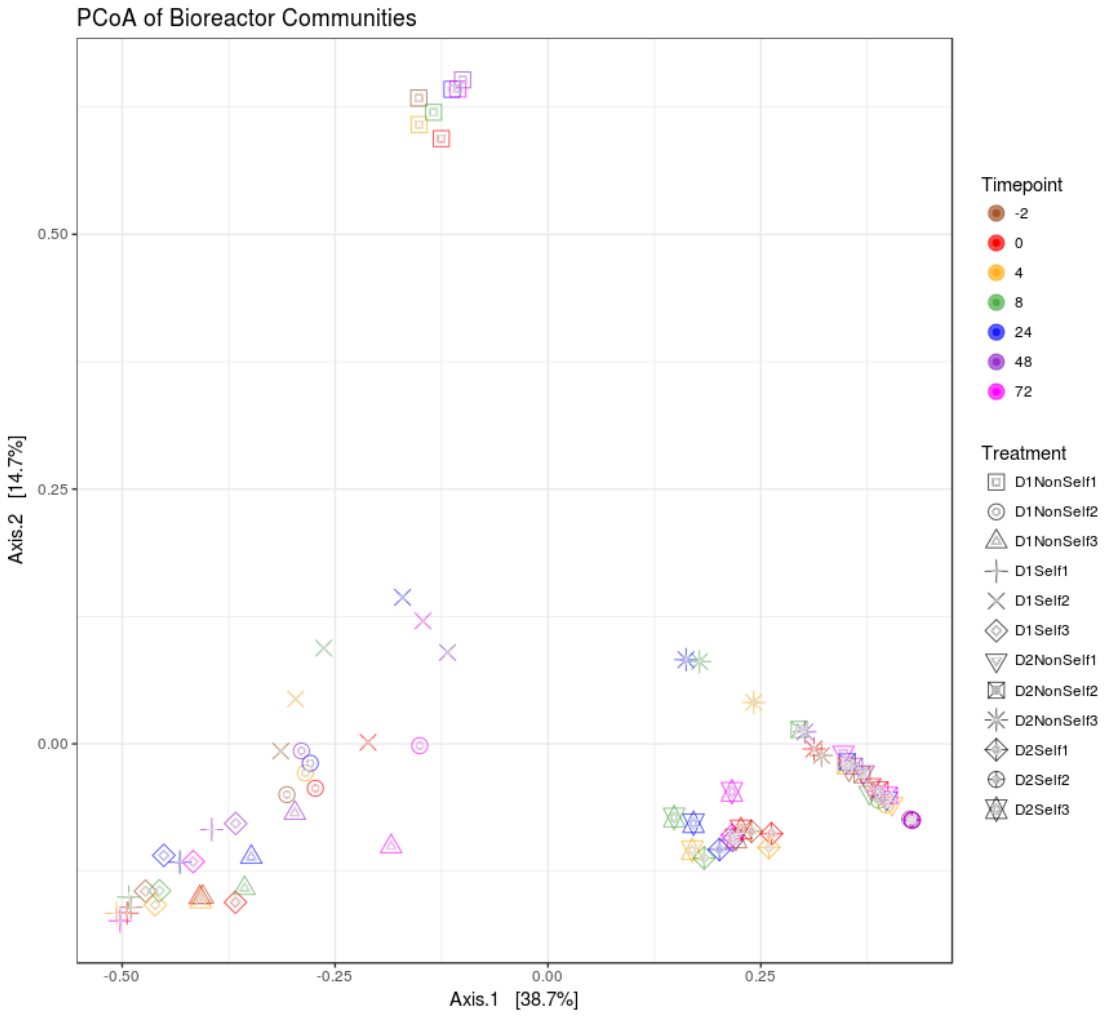

Figure 12: MBRA Principle Coordinate Analysis. PCoA of all samples coded by reactor (shape) and time point (color). Ordination based on Bray-Curtis dissimilarity. The first major axis depicts inter-donor variation while the second major axis captures the noted variation within Donor 1 reactors.

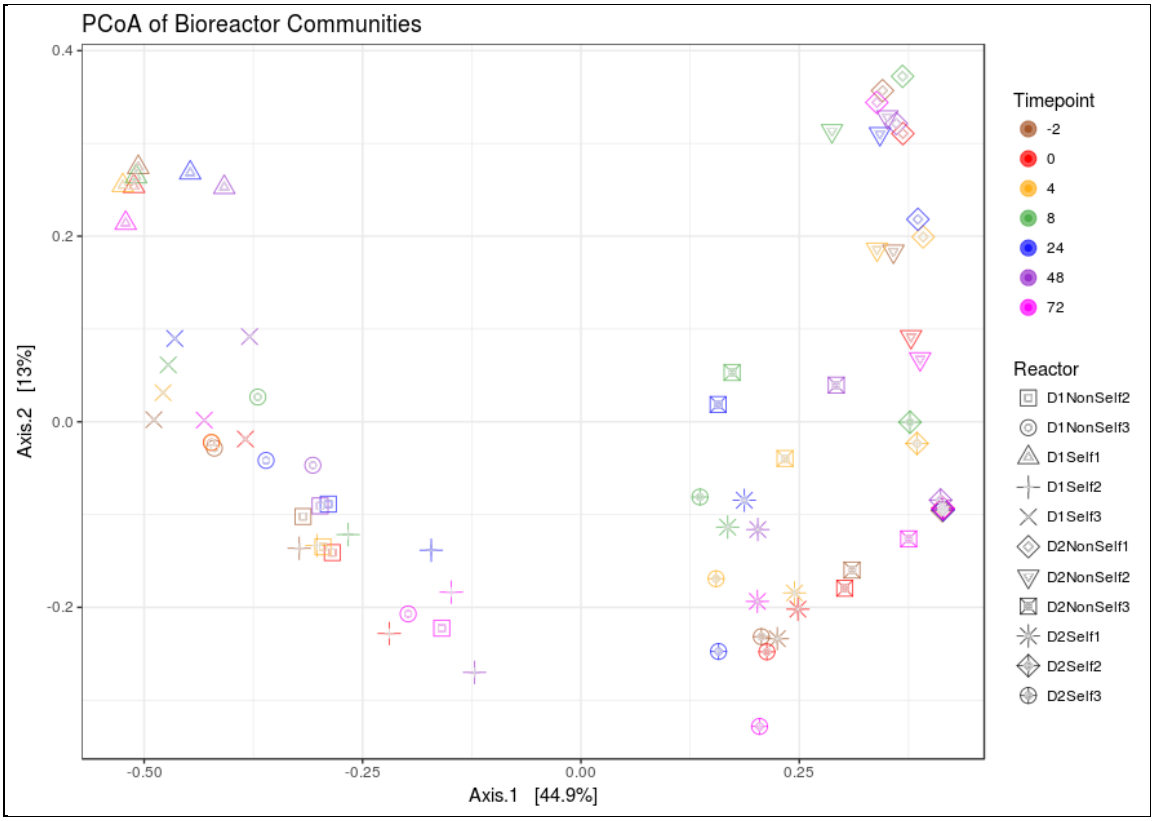

Figure 13: MBRA Principle Coordinate Analysis – outlier removed. PCoA of all samples except Donor 1 outlier reactor samples. Samples are differentiated by reactor (shape) and time point (color) and ordination distance is based on Bray-Curtis dissimilarity index.

Emergent trends in community fluctuation between the treatment groups within donor strata were uncovered by performing PCoA on individual treatment groups (Figure 14). By limiting the scope of principle coordinates to reactors related by treatment and donor, any common trends at the community level will manifest as directional changes along one or multiple axes. For all four PCoA's the greatest source of variation, represented in axis 1, was primarily composed of inter-reactor variation. A uniform shift occurred in both Donor 1 and 2 autologous treatment groups. The common directional shift coded in axis 2 was modeled against time using a quadratic linear regression which produced significant associations in both donors. These same time dependent trends were

not apparent within any of the major principle coordinates for the non-autologous treated reactors of either donor and no significant models were generated.

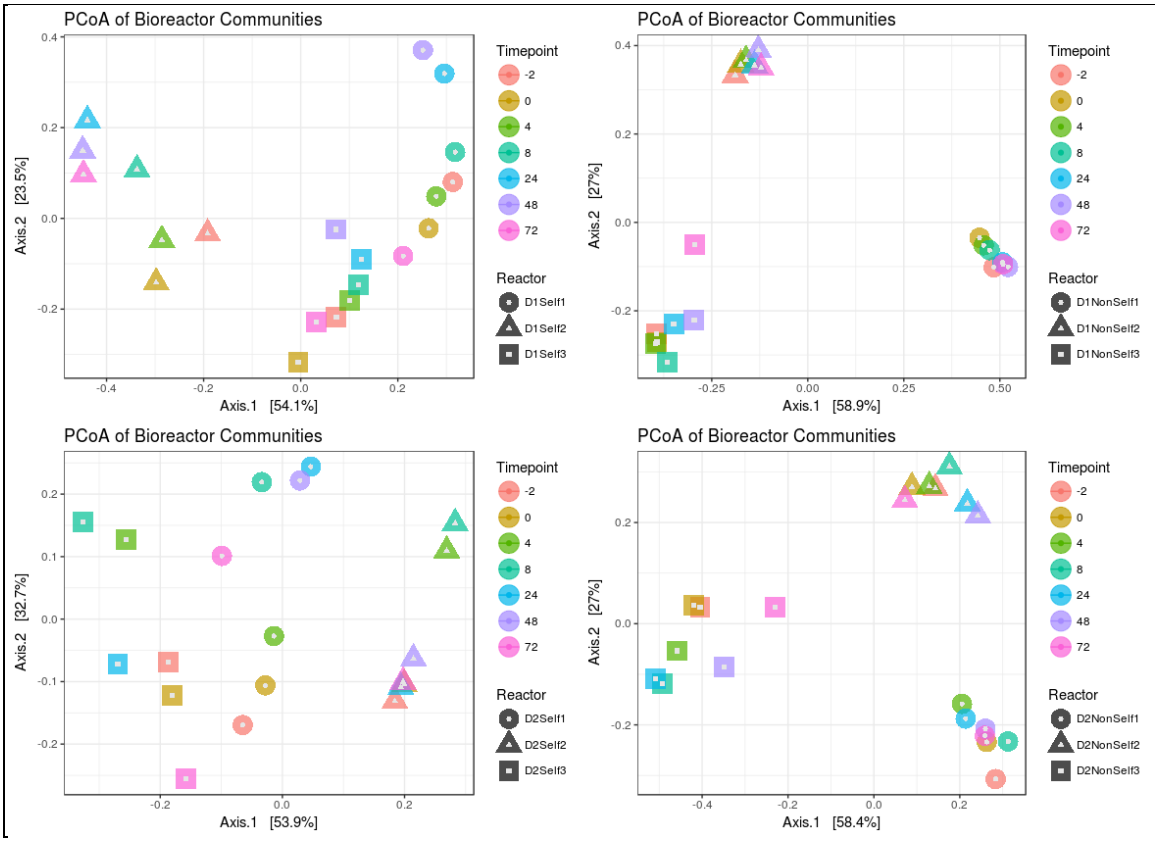

Figure 14: PCoA of individual treatment groups. Top panels represent Donor 1 reactors with self treated reactors on the left and non-self treated reactors ordinate on the right. The lower panels represent Donor 2 reactors. Ordination shapes depict specific reactors and are colored by the time point with "0" referencing the time of phage spike.

Discussion

Nutrient niche theory has been demonstrated to control the cellular and in turn viral communities of the GI microbiota (100). Though perhaps intuitive, the inverse of this concept has not been conclusively illustrated with respect to viruses modulating host associated microbial communities. In this study, alterations in 16S rRNA community profiles likely resulted from top down disturbance by viruses and by possibly other parasitic elements.

Bioreactor Communities

High representation of Firmicutes taxa in this16S rRNA amplicon dataset should be considered with the caveat that the V4 region preferentially amplifies the Firmicutes and can underrepresent Bacteroides abundances up to 4-fold from the true community composition (Ibarbalz 2014). As noted in previous Britton lab publications, the MBRA systems standard media formulation also enrich for members of the Firmicutes phylum (68, 69).

Community profiles at the phylum level also illustrate potential cluster effects. The same coordinate reactor for each donor contained a considerable proportion of Bacteroidetes taxa. The inoculation procedure may have caused heterogeneity across reactors despite slurry material coming from a common vial. However, at this stage, these are mere speculations and a certain degree of heterogeneity is expected from a complex human fecal environmental sample. Unfortunately, input fecal samples used to initiate the MBRAs were not sequenced. In previous work, Dr. Britton's group reported

incompatibility in sequencing outputs between fecal and MBRA samples (68). The physical and chemical composition of the two sample types do not lend themselves to comparable DNA extraction efficiencies across bacterial taxonomies.

Viral Dynamics

The three MGEs tested display three distinct processes among many possible modes. It is reassuring that within three somewhat arbitrarily selected targets, two show consistent trends across replicate reactors. Titers did drop over time for the tracked viruses. They were able to persist up to the final sampling point in some instances, however. At the titers measured, they would almost certainly have had to replicate in the reactors where they remained detectable. These data support the relevance of complex continuous culture platforms for studying the interaction of prokaryotic viruses and their hosts. In future experiments, analysis at the viral community level using metagenomics would likely contribute a great deal of insight beyond what we were able to track quantitatively.

Bacterial Dynamics

Individuals harbor gut microbial communities which have unique structural profiles. In other words, at a given time the microbiota of a person will retain compositional hallmarks that more closely resemble the community composition of that same person at other times than the community composition of a different individual. Ideally, this property will be maintained within the bioreactors. The analyzed data supports this notion. Individual reactors harbored unique microbial assemblies even within donor sets. This indicates that from inoculation they begin to differentiate from

their corresponding replicates while maintaining aspects of the donor community. Inferences regarding whole community composition are possible based on these calculations, however co-variation of communities or community components with environmental and imposed variables such as time and phage treatment are beyond the scope of PERMANOVA and beta-dispersion analysis.

The differential responses of bacterial communities to the two types of viral filtrate were consistent across donors and reactors. The mostly uniform directional shifts were apparent in the autologous treated reactors as evident when the ordinations were limited to those within the autologous treatment group of either donor. This strategy served as a proxy for constrained ordination without relying on environmental variables. Imposing the time variable against PCoA axis 2 is synonymous with plotting against the phage challenge effect through time since the two variables are collinear. Interpretation of this correlation is limited, however, we can safely say that the secondary axis for autologous PCoA's encodes for community level variation corresponding to the phage/time variables and is consistent across replicate reactors and the two donor communities tested. This same effect was not readily apparent in non-autologous treated reactors in any of the major ordination axes. Though effects may have been present, they were not reproduced across replicates to a degree greater than the stochastic variability within reactors or the inter-replicate variability. It should be noted that the outlying reactor of the Donor 1 set does complicate this interpretation for the Donor 1 non-autologous ordination since its 'outlying' effect causes reactor groupings to tighten. This

could reasonably mask any time/phage correlated effects however omission of the outlying samples does not render a tractable effect.

Conclusion

Bacterial communities clearly responded to extracellular filtrates enriched for bacteriophage, supporting the hypothesis that extracellular elements, especially phage, can impact the cellular community structure of the human GI microbiota . The results of these experiments support the conclusions that phage are involved in top down regulation of community assembly and that autologous and non-autologous phage differ in their capacity to interact with cellular communities. Autologous derived phage elicited a structured, reproducible effect and non-autologous phage generated a more stochastic and less tractable change. Although bacterial community complexity was considerably different between the two donors, this dimension of composition did not noticeably alter the effect of viral filtrates. A combination of inter-individual variability and localized adaptation of shared viral taxa are logical sources of the observed community responses and would be reasonable subjects for future investigation. Overall, the MBRA environment provided a stable nutrient landscape for complex microbiota which in turn served as the niche habitat for viruses and other MGEs

It is important to note several limitations of these experiments and their interpretation. The lack of untreated controls does not allow much latitude for interpretations. Second, the observed community balance is a potential concern with the

noted overrepresentation of Firmicutes and possible cluster effects within donor groups. Future experiments should be designed to overcome these potential limitations.

Bacteriophage, identified with metagenomics, were able to replicate within the system. The two putative viruses demonstrated unique establishment and maintenance regimes. These dynamics were surprisingly consistent across replicate reactors. Despite inter-reactor variation, a host landscape was present for the individual phage tested. Initial drops in titer during the reactor establishment phage and subsequent ability of those same phage to colonize following phage spike may present evidence of co-evolutionary dynamics. Alternatively, the pervasive nature of these particular phage in the human population may reflect a generalist strategy. Expanding this analysis with viral metagenomics would likely yield important insights to mechanistic interactions of phages with their cellular hosts.

Scope of Inference

This analysis investigated basic structuring of bacterial community data from an *in-vitro* experiment with replication and temporal sampling as well as quantitative tracking of a subset of viruses and other MGE's. Inferences are limited to this specific dataset with the potential to inform expectations for future experiments using similar methods.

CHAPTER FOUR

PERSPECTIVES AND FUTURE DIRECTIONS

Findings

This thesis sought to simplify the vastly complicated human ecosystem to two crucial and inexorably linked components within the GI tract: the gut microbial community and its viruses. We applied a top-down systems approach to isolate the GI microbiota of healthy humans and empirically test the effect of extracellular elements on community structure. The experiments carried out for this thesis were designed to test the following two hypotheses and addressed two main themes. First, applicability of MBRA continuous cultures was investigated using observation based inquiry to validate the experimental platform. This thrust carried across to the experimental investigation itself of fundamental phage ecology; the effect(s) of viruses on gut microbial community structure . Both the application and fundamental inquiries yielded interpretable findings.

Hypothesis 1:

MBRA based continuous culture systems seeded with human fecal microbiota harbor considerable bacteriophage diversity including phage taxonomies commonly found in healthy humans.

Hypothesis 2:

Extracellular elements, especially phage, impact the cellular community structure of the human GI microbiota.

Application Based Results:

One advantage of conducting ecological surveys at the microbial scale is the ability to thoroughly and reproducibly interrogate entire biomes. Next generation sequencing and other 'omic' techniques have bolstered this notion and provided salient advances in our understanding of microbial habitats. However, very little has been done experimentally to understand the fundamental ecology of phage and other mobile genetic elements at a community level. The results reported in this work begin to move us beyond correlative studies and toward an ecological understanding of viruses within the human GI microbiome. In developing this thesis we asked: Can we empirically study host associated phage ecology in a top-down manner using the MBRA continuous culture system?

Minibioreactors supported a diverse phage community similar to what was observed in larger scale chemostat systems. Phage communities from both Santiago-Rodriguez et al. and MBRA samples both reliably contained genetic signatures of highly conserved phage types. We observed a considerable overlap between our MSU donors' viromes and those found in both mined and newly generated data from MBRAs humanized with unrelated donor material.

Using analytical techniques from quantitative genome tracking to community and amplicon-based metagenomics we were able to track individual MGE dynamics as well as bacterial community responses to different extracellular assemblies. Although individual bioreactors exhibit properties somewhere between technical and biological replicates, a high degree of reproducibility was observed in the dynamics of two of the

three screened MGE's. Bacterial communities also responded consistently within individual MBRAs depending on extracellular treatment.

In-vitro continuous culture systems can support a diversity of previously un-isolated bacteria and viral taxa. Continuous flow conditions allow dynamic interactions between microbial cellular communities and viruses to be tested. These findings support the first hypothesis and use of bioreactors for investigation of the human GI microbiome.

Fundamental Results:

From daily life to the extreme event of fecal transplantation, humans constantly encounter exogenous microbial elements. Through this ebb and flow, individual communities retain surprising interpersonal variation. Observational studies indicate that microbial viruses are not only part of a homeostatic host ecosystem but a major element within that system. Phage and other MGE's impact individual microbial populations along a spectrum of known interactions and likely many more yet to be characterized. Previous work has examined the impact of pathogen specific phage therapy on particular deleterious microbes (101). However, at the community and ecosystem level many questions remain scantly formulated. We investigated the ability of phage and other parasitic elements to affect GI microbial community structure when established in the MBRA system.

We found evidence of the establishment of an exogenous virus applied to a gut microbial community as well as potential variant competition within another phage species. At the cellular level we observed two distinct modes of community response, structured and stochastic, to challenge with either self or non-self viral filtrates

respectively. The results directly support hypothesis two. These findings may offer insight to the maintenance of stable community assembly dynamics that vary between individuals, diet, and specific health state.

Summary of Techniques: The ‘ome’, the whole ‘ome’, and nothing but the ‘ome’?

Early in the course of developing this project we contemplated a number of experimental techniques. The first plan was to develop a pathogen free 15 species bacterial consortium for highly controlled community investigations. Integrated prophage were to be disabled via targeted genome editing. Gnotobiotic animal models and *in-vitro* co-culture would then be used to investigate community structure with and without active phage. One can imagine the many advantages to exercising this level of control over an experimental system. Despite the advantages, it was difficult to reconcile the highly contrived nature of this approach with the complexity of the real human GI microbiome. As discussed in Chapter 1, the microbiome has a high degree of strain diversity and functional redundancy. Accurately testing the holistic role of phage within this system likely requires us to recapitulate any emergent properties related to its complexity. We decided on a less controlled, higher complexity system aided by cellular and viral metagenomics. Thus, in a sense we gained a certain degree of control by removing confounding aspects of the original environment while maintaining enough unknown elements of the community to make more relatable observations.

Omic Advantages

Quantitative tracking of individual MGE's allowed us to chart the fate of these elements throughout the experimental time course. While we were able to address certain applications based questions and infer certain dynamics, this data does little to relate to the system as a whole. Scaling of qPCR to the community level is unrealistic. Microarray technology could be useful but successfully capturing the interpersonal variation beyond one or two donors is also not feasible. Here metagenomics truly shines in its ability to provide nearly qPCR level detection of a huge diversity of genetic elements across donor groups. Appropriately deep metagenomes can also provide a wealth of information regarding the pangenome or microdiversity of viruses and other extracellular elements.

Targeted cellular methods are not easily scalable to the community level. Some scaling is possible with corresponding losses of resolution in that higher taxonomic groupings can be tracked up to the phyla and kingdom level but provide no insights to lower taxonomic level differences . However, quantifying different phylogenetic levels can be useful for benchmarking inter- and intra-sample metagenomic sequencing.

We utilized amplicon based metagenomics as an affordable method to track bacterial communities. This method provided near species level resolution with reasonable computational requirements. Considerably more samples were able to be analyzed at a fraction of the expense of complete cellular metagenomes owing to lower library preparation cost and lower sequencing depth requirements.

Omic Disadvantages

Metagenomics offer exponentially more community information than targeted methods with some key disadvantages. Unlike a readily repeated, targeted assay, deep sequencing is not easily reproduced due to monetary and time expenditure. NGS processes are also extremely susceptible to contamination and can suffer from within run bleed-over of samples (102). Computational demands are dramatically higher when analyzing NGS datasets. Depending on experimental design, results can be difficult to interpret where isolated molecular studies are generally clear but can be difficult to extrapolate.

Future Directions

Future work will ideally focus on testing the MBRA gut virus-microbe community model more extensively. Along with adding inactivated virus and non-treated controls, the analysis should be repeated using at least one separate donor pair. Heat killed or ultra filtered supernatants are good candidates for control treatments. Both serve to disable or remove viruses and vesicle encapsulated elements. Heating, more accurately, boiling is less prone to malfunction but does alter the chemical matrix of the filtrate in a considerable way.

Technically speaking, microbiota cultivation could be improved by amending media with relevant elements like mucin. In their 2015 paper, Dr. Britton's group discusses possible strategies to encourage the growth of *Faecalibacterium prausnitzii*, an abundant microbiome member, within reactor communities (68). Though procedurally

difficult, randomization of treatment groups across reactor arrays would alleviate any concerns regarding cluster effects.

DNA Approaches

Future investigations will be able to employ directed allocation of sequencing resources. Coordination between cellular and viral sequencing could provide a windfall of data regarding dynamic interactions of the extracellular and cellular communities during perturbation or challenge. Co-evolutionary interactions could potentially be observed using this strategy.

Complete cellular metagenomes would expand 16S rRNA amplicon findings to other kingdoms and pangenomic/sub-species levels within bacteria. However, the additional cost and computational burden associated with whole cellular metagenomics is difficult to justify (103). Without clear justification it is probably more advantageous to invest resources in improved amplicon approaches. Paired microfluidic-amplicon approaches could offer advantages over single hypervariable amplicon analysis (104, 105).

Alternative Techniques

Other approaches can be leveraged for the study of phage-host interaction using the MBRA system. These can be used in combination with, or in addition to, DNA analysis. Physical partitioning of complex samples based on dilution or fluorescent tagging is a rapidly developing practice for interrogating microbial community processes (106-114). These strategies could be adapted to the MBRA study system to ascertain

things like virus host range. Work in marine environments has shown that free ribosomes can remain stable in the extracellular environment and serve as a marker of population dynamics and potentially phage-host interaction (Unpublished data Dr. Curtis Suttle's research group University of British Columbia). Combining reverse transcription of free-ribosomes in bioreactor effluent with standard community 16S rRNA amplicon profiling and paired viral metagenomics could allow us to infer viral-host coupled processes. Isolation based analysis can be used to investigate lytic and lysogenic evolutionary interactions (115).

Paradigms

Co-culturing allows us to move beyond *in-situ* observational constraints to address questions experimentally at the community level. The MBRA system was able to maintain microbial taxonomies which have not previously been isolated along with a diversity of uncharacterized phage likely infecting them. The extracellular cross results illustrate a potential archetype for community structuring within the GI microbiome. Firstly, phage are involved in top-down community regulation and can thereby contribute to symbiosis within metazoan hosts. Also, community level susceptibility to phage effects is origin specific. Local adaptation of phage communities may be functionally important in addition to inter-personal variation in community composition. Deeper understanding of viruses in the human microbiome is highly relevant to understanding and developing microbiome based medicine. Where current practices such as fecal transplantation introduce a litany of microbial elements to varying effect, targeted

intervention with selected elements of the microbiome, such as viruses, would offer more rigorous precision which modern medicine demands.

Where quintessential disease models paint viruses and parasitic elements as deleterious agents, we are beginning to appreciate their other roles in community homeostasis. Parallelized MBRA based co-culture is a useful approach for gaining spatial and temporal resolution to complex communities from the human GI tract. In combination with clinical studies and other model systems it should be useful in disentangling the ecological function of viruses and other MGEs. As our results indicate, this requires defining abstract concepts of health and homeostasis in a whole ecosystem context.

REFERENCES CITED

1. Petersen C & Round JL (2014) Defining dysbiosis and its influence on host immunity and disease. *Cell Microbiol* 16(7):1024-1033.

2. Ehrlich SD & Consortium M (2011) MetaHIT: The European Union Project on Metagenomics of the Human Intestinal Tract. *Metagenomics of the Human Body*:307-316.

3. Human Microbiome Project C (2012) Structure, function and diversity of the healthy human microbiome. *Nature* 486(7402):207-214.

4. Turnbaugh PJ, *et al.* (2007) The human microbiome project. *Nature* 449(7164):804-810.

5. Costello EK, *et al.* (2009) Bacterial community variation in human body habitats across space and time. *Science* 326(5960):1694-1697.

6. Flores GE, *et al.* (2014) Temporal variability is a personalized feature of the human microbiome. *Genome Biology* 15(12).

7. Backhed F, *et al.* (2012) Defining a healthy human gut microbiome: current concepts, future directions, and clinical applications. *Cell Host Microbe* 12(5):611-622.

8. Yatsunenko T, *et al.* (2012) Human gut microbiome viewed across age and geography. *Nature* 486(7402):222-227.

9. Sender R, Fuchs S, & Milo R (2016) Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biol* 14(8):e1002533.

10. Faith JJ, *et al.* (2013) The long-term stability of the human gut microbiota. *Science* 341(6141):1237439.

11. Greenblum S, Carr R, & Borenstein E (2015) Extensive strain-level copy-number variation across human gut microbiome species. *Cell* 160(4):583-594.

12. Ley RE, *et al.* (2008) Evolution of mammals and their gut microbes. *Science* 320(5883):1647-1651.

13. Schloissnig S, *et al.* (2013) Genomic variation landscape of the human gut microbiome. *Nature* 493(7430):45-50.

14. Zhu A, Sunagawa S, Mende DR, & Bork P (2015) Inter-individual differences in the gene content of human gut bacterial species. *Genome Biology* 16.

15. Franks AH, *et al.* (1998) Variations of bacterial populations in human feces measured by fluorescent in situ hybridization with group-specific 16S rRNA-targeted oligonucleotide probes. *Appl Environ Microbiol* 64(9):3336-3345.

16. Hoyles L, *et al.* (2014) Characterization of virus-like particles associated with the human faecal and caecal microbiota. *Res Microbiol* 165(10):803-812.

17. Stephen AM & Cummings JH (1980) The microbial contribution to human faecal mass. *J Med Microbiol* 13(1):45-56.

18. Kraal L, Abubucker S, Kota K, Fischbach MA, & Mitreva M (2014) The prevalence of species and strains in the human microbiome: a resource for experimental efforts. *PLoS One* 9(5):e97279.

19. Clemente JC, Ursell LK, Parfrey LW, & Knight R (2012) The impact of the gut microbiota on human health: an integrative view. *Cell* 148(6):1258-1270.

20. Lloyd-Price J, Abu-Ali G, & Huttenhower C (2016) The healthy human microbiome. *Genome Med* 8(1):51.

21. Norman JM, *et al.* (2015) Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell* 160(3):447-460.

22. Perez-Brocal V, *et al.* (2013) Study of the viral and microbial communities associated with Crohn's disease: a metagenomic approach. *Clin Transl Gastroenterol* 4:e36.

23. Aguirre M, Bussolo de Souza C, & Venema K (2016) The Gut Microbiota from Lean and Obese Subjects Contribute Differently to the Fermentation of Arabinogalactan and Inulin. *PLoS One* 11(7):e0159236.

24. Ridaura VK, *et al.* (2013) Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science* 341(6150):1241214.

25. Clarke SF, *et al.* (2012) The gut microbiota and its relationship to diet and obesity: new insights. *Gut Microbes* 3(3):186-202.

26. Ley RE, *et al.* (2005) Obesity alters gut microbial ecology. *Proc Natl Acad Sci U S A* 102(31):11070-11075.

27. Ley RE, Turnbaugh PJ, Klein S, & Gordon JI (2006) Microbial ecology: human gut microbes associated with obesity. *Nature* 444(7122):1022-1023.

28. Turnbaugh PJ, *et al.* (2009) A core gut microbiome in obese and lean twins. *Nature* 457(7228):480-484.

29. Walters WA, Xu Z, & Knight R (2014) Meta-analyses of human gut microbes associated with obesity and IBD. *FEBS Lett* 588(22):4223-4233.

30. Shin NR, Whon TW, & Bae JW (2015) Proteobacteria: microbial signature of dysbiosis in gut microbiota. *Trends Biotechnol* 33(9):496-503.

31. Arumugam M, *et al.* (2011) Enterotypes of the human gut microbiome. *Nature* 473(7346):174-180.

32. Jeffery IB, Claesson MJ, O'Toole PW, & Shanahan F (2012) Categorization of the gut microbiota: enterotypes or gradients? *Nat Rev Microbiol* 10(9):591-592.

33. Knights D, *et al.* (2014) Rethinking "enterotypes". *Cell Host Microbe* 16(4):433-437.

34. Qin J, *et al.* (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464(7285):59-65.

35. Kim MS, Park EJ, Roh SW, & Bae JW (2011) Diversity and abundance of single-stranded DNA viruses in human feces. *Appl Environ Microbiol* 77(22):8062-8070.

36. Lepage P, *et al.* (2008) Dysbiosis in inflammatory bowel disease: a role for bacteriophages? *Gut* 57(3):424-425.

37. Breitbart M, *et al.* (2003) Metagenomic analyses of an uncultured viral community from human feces. *J Bacteriol* 185(20):6220-6223.

38. Reyes A, *et al.* (2010) Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* 466(7304):334-U381.

39. Santiago-Rodriguez TM, *et al.* (2015) Chemostat culture systems support diverse bacteriophage communities from human feces. *Microbiome* 3:58.

40. Minot S, *et al.* (2011) The human gut virome: inter-individual variation and dynamic response to diet. *Genome Res* 21(10):1616-1625.

41. Zhang T, *et al.* (2006) RNA viral community in human feces: prevalence of plant pathogenic viruses. *PLoS Biol* 4(1):e3.

42. Castro-Mejia JL, *et al.* (2015) Optimizing protocols for extraction of bacteriophages prior to metagenomic analyses of phage communities in the human gut. *Microbiome* 3.

43. Reyes A, *et al.* (2015) Gut DNA viromes of Malawian twins discordant for severe acute malnutrition. *Proc Natl Acad Sci U S A* 112(38):11941-11946.

44. Minot S, *et al.* (2013) Rapid evolution of the human gut virome. *Proc Natl Acad Sci U S A* 110(30):12450-12455.

45. Dutilh BE, *et al.* (2014) A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nature Communications* 5.

46. Manrique P, *et al.* (2016) Healthy human gut phageome. *Proc Natl Acad Sci U S A* 113(37):10400-10405.

47. Paez-Espino D, *et al.* (2016) Uncovering Earth's virome. *Nature* 536(7617):425-430.

48. 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(10):1985-1994.

49. Modi SR, Lee HH, Spina CS, & Collins JJ (2013) Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome. *Nature* 499(7457):219-222.

50. Monaco CL, *et al.* (2016) Altered Virome and Bacterial Microbiome in Human Immunodeficiency Virus-Associated Acquired Immunodeficiency Syndrome. *Cell Host Microbe* 19(3):311-322.

51. Geier MR, Trigg ME, & Merril CR (1973) Fate of bacteriophage lambda in non-immune germ-free mice. *Nature* 246(5430):221-223.

52. Keller R & Engley FB (1958) Fate of Bacteriophage Particles Introduced into Mice by Various Routes. *P Soc Exp Biol Med* 98(3):577-580.

53. Bahra SM, *et al.* (2015) Risperidone-induced weight gain is mediated through shifts in the gut microbiome and suppression of energy expenditure. *EBioMedicine* 2(11):1725-1734.

54. 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(8):e0134941.

55. Ly M, *et al.* (2016) Transmission of viruses via our microbiomes. *Microbiome* 4.

56. 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(50):20236-20241.

57. Lotka AJ (1920) Analytical note on certain rhythmic relations in organic systems. *P Natl Acad Sci USA* 6:410-415.

58. Volterra V (1927) Fluctuations in the abundance of a species considered mathematically. *Nature* 119:12-13.

59. De Paepe M, *et al.* (2016) Carriage of. Latent Virus Is Costly for Its Bacterial Host due to Frequent Reactivation in Monoxenic Mouse Intestine. *Plos Genetics* 12(2).

60. Duerkop BA, Clements CV, Rollins D, Rodrigues JL, & Hooper LV (2012) A composite bacteriophage alters colonization by an intestinal commensal bacterium. *Proc Natl Acad Sci U S A* 109(43):17621-17626.

61. Kim HJ & Ingber DE (2013) Gut-on-a-Chip microenvironment induces human intestinal cells to undergo villus differentiation. *Integr Biol (Camb)* 5(9):1130-1140.

62. Barr JJ, *et al.* (2013) Bacteriophage adhering to mucus provide a non-host-derived immunity. *Proc Natl Acad Sci U S A* 110(26):10771-10776.

63. Barr JJ, *et al.* (2015) Subdiffusive motion of bacteriophage in mucosal surfaces increases the frequency of bacterial encounters. *Proc Natl Acad Sci U S A* 112(44):13675-13680.

64. Nguyen S, *et al.* (2017) Bacteriophage Transcytosis Provides a Mechanism To Cross Epithelial Cell Layers. *MBio* 8(6).

65. D'Herelle F (2007) On an invisible microbe antagonistic toward dysenteric bacilli: brief note by Mr. F. D'Herelle, presented by Mr. Roux. 1917. *Res Microbiol* 158(7):553-554.

66. Twort FW (1915) An investigation on the nature of ultra-microscopic viruses. *Lancet* 2:1241-1243.

67. Chibani-Chennoufi S, *et al.* (2004) In vitro and in vivo bacteriolytic activities of Escherichia coli phages: implications for phage therapy. *Antimicrob Agents Chemother* 48(7):2558-2569.

68. Auchtung JM, Robinson CD, & Britton RA (2015) Cultivation of stable, reproducible microbial communities from different fecal donors using minibioreactor arrays (MBRAs). *Microbiome* 3.

69. Auchtung JM, Robinson CD, Farrell K, & Britton RA (2016) MiniBioReactor Arrays (MBRAs) as a Tool for Studying C. difficile Physiology in the Presence of a Complex Community. *Methods Mol Biol* 1476:235-258.

70. Kim BS, Kim JN, & Cerniglia CE (2011) In vitro culture conditions for maintaining a complex population of human gastrointestinal tract microbiota. *J Biomed Biotechnol* 2011:838040.

71. McDonald JA, *et al.* (2013) Evaluation of microbial community reproducibility, stability and composition in a human distal gut chemostat model. *J Microbiol Methods* 95(2):167-174.

72. Rajilic-Stojanovic M, *et al.* (2010) Evaluating the microbial diversity of an in vitro model of the human large intestine by phylogenetic microarray analysis. *Microbiology* 156(Pt 11):3270-3281.

73. Van den Abbeele P, *et al.* (2010) Microbial Community Development in a Dynamic Gut Model Is Reproducible, Colon Region Specific, and Selective for Bacteroidetes and Clostridium Cluster IX. *Appl Environ Microb* 76(15):5237-5246.

74. Kaur A, Rose DJ, Rumpagaporn P, Patterson JA, & Hamaker BR (2011) In vitro batch fecal fermentation comparison of gas and short-chain fatty acid production using "slowly fermentable" dietary fibers. *J Food Sci* 76(5):H137-142.

75. Khalil NA, Walton GE, Gibson GR, Tuohy KM, & Andrews SC (2014) In vitro batch cultures of gut microbiota from healthy and ulcerative colitis (UC) subjects suggest that sulphate-reducing bacteria levels are raised in UC and by a protein-rich diet. *Int J Food Sci Nutr* 65(1):79-88.

76. Macfarlane GT & Macfarlane S (2007) Models for intestinal fermentation: association between food components, delivery systems, bioavailability and functional interactions in the gut. *Curr Opin Biotechnol* 18(2):156-162.

77. Mirzaei MK & Maurice CF (2017) Menage a trois in the human gut: interactions between host, bacteria and phages. *Nature Reviews Microbiology* 15(7):397-408.

78. Jover LF, Romberg J, & Weitz JS (2016) Inferring phage-bacteria infection networks from time-series data. *R Soc Open Sci* 3(11):160654.

79. Ahlgren NA, Ren J, Lu YY, Fuhrman JA, & Sun F (2017) Alignment-free d_2^* oligonucleotide frequency dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences. *Nucleic Acids Res* 45(1):39-53.

80. Jonczyk E, Klak M, Miedzybrodzki R, & Gorski A (2011) The influence of external factors on bacteriophages--review. *Folia Microbiol (Praha)* 56(3):191-200.

81. Santiago-Rodriguez TM, Ly M, Bonilla N, & Pride DT (2015) The human urine virome in association with urinary tract infections. *Front Microbiol* 6:14.

82. Meyer F, *et al.* (2008) The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9:386.

83. Peng Y, Leung HC, Yiu SM, & Chin FY (2012) IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28(11):1420-1428.

84. Kearse M, *et al.* (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12):1647-1649.

85. Langmead B & Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357-359.

86. Yutin N, *et al.* (2018) Discovery of an expansive bacteriophage family that includes the most abundant viruses from the human gut. *Nature Microbiology* 3(1).

87. Robinson CD, Auchtung JM, Collins J, & Britton RA (2014) Epidemic Clostridium difficile strains demonstrate increased competitive fitness compared to nonepidemic isolates. *Infect Immun* 82(7):2815-2825.

88. Wu S, Zhu Z, Fu L, Niu B, & Li W (2011) WebMGA: a customizable web server for fast metagenomic sequence analysis. *BMC Genomics* 12:444.

89. Kristensen DM, *et al.* (2013) Orthologous gene clusters and taxon signature genes for viruses of prokaryotes. *J Bacteriol* 195(5):941-950.

90. Altschul SF, *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25(17):3389-3402.

91. Quast C, *et al.* (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41(Database issue):D590-596.

92. Altschul SF, Gish W, Miller W, Myers EW, & Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403-410.

93. Ott SJ, *et al.* (2016) Efficacy of Sterile Fecal Filtrate Transfer for Treating Patients With Clostridium Difficile Infection. *Gastroenterology*.

94. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, & Schloss PD (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* 79(17):5112-5120.

95. Jari Oksanen FGB, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens,Eduard Szoecs, Helene Wagner (2017) vegan: Community Ecology Package. R package version 2.4-4.https://CRAN.R-project.org/package=vegan.

96. McMurdie PJ & Holmes S (2013) phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8(4):e61217.

97. RStudio Team (2015) RStudio: Integrated Develpment for R. in *RStudio, Inc., Boston, MA.*

98. Westcott SL & Schloss PD (2017) OptiClust, an Improved Method for Assigning Amplicon-Based Sequence Data to Operational Taxonomic Units. *mSphere* 2(2).

99. Anderson MJ (2001) A new method for non-parametric multivariate analysis of variance. *Austral Ecol* 26(1):32-46.

100. Pereira FC & Berry D (2017) Microbial nutrient niches in the gut. *Environ Microbiol* 19(4):1366-1378.

101. Abedon ST, Kuhl SJ, Blasdel BG, & Kutter EM (2011) Phage treatment of human infections. *Bacteriophage* 1(2):66-85.

102. Wright ES & Vetsigian KH (2016) Quality filtering of Illumina index reads mitigates sample cross-talk. *BMC Genomics* 17(1):876.

103. Segata N, *et al.* (2012) Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Methods* 9(8):811-814.

104. Caro-Quintero A & Ochman H (2015) Assessing the Unseen Bacterial Diversity in Microbial Communities. *Genome Biol Evol* 7(12):3416-3425.

105. Hermann-Bank ML, Skovgaard K, Stockmarr A, Larsen N, & Molbak L (2013) The Gut Microbiotassay: a high-throughput qPCR approach combinable with next generation sequencing to study gut microbial diversity. *BMC Genomics* 14:788.

106. Allen LZ, *et al.* (2011) Single Virus Genomics: A New Tool for Virus Discovery. *Plos One* 6(3).

107. Allen LZ, *et al.* (2013) Isolation and Genome Analysis of Single Virions using 'Single Virus Genomics'. *Jove-J Vis Exp* (75).

108. Hatzenpichler R, *et al.* (2016) Visualizing in situ translational activity for identifying and sorting slow-growing archaeal-bacterial consortia. *Proc Natl Acad Sci U S A* 113(28):E4069-4078.

109. Martinez-Hernandez F, *et al.* (2017) Single-virus genomics reveals hidden cosmopolitan and abundant viruses. *Nat Commun* 8:15892.

110. Martinez Martinez J, Swan BK, & Wilson WH (2014) Marine viruses, a genetic reservoir revealed by targeted viromics. *Isme J* 8(5):1079-1088.

111. Porter J, Robinson J, Pickup R, & Edwards C (1995) Recovery of a Bacterial Subpopulation from Sewage Using Immunofluorescent Flow-Cytometry and Cell Sorting. *Fems Microbiol Lett* 133(1-2):195-199.

112. Tadmor AD, Ottesen EA, Leadbetter JR, & Phillips R (2011) Probing Individual Environmental Bacteria for Viruses by Using Microfluidic Digital PCR. *Science* 333(6038):58-62.

113. Roux S, *et al.* (2014) Ecology and evolution of viruses infecting uncultivated SUP05 bacteria as revealed by single-cell- and meta- genomics. *Elife* 3.

114. Labonte JM, *et al.* (2015) Single-cell genomics-based analysis of virus-host interactions in marine surface bacterioplankton. *Isme J* 9(11):2386-2399.

115. Koskella B & Brockhurst MA (2014) Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *Fems Microbiol Rev* 38(5):916-931.

APPENDICES

APPENDIX A

MODIFIED MOTHUR SOP

mothur > make.contigs(ffastq=BR67_R1.fastq, rfastq=BR67_R2.fastq,
findex=BR67_I1.fastq, rindex=BR67_I2.fastq, oligos=BR67_oligos_formatted.oligos,
checkorient=T, processors=32)

mothur > system(mv BR67_R1.trim.contigs.fasta BR67.trim.contigs.fasta)

mothur > system(mv BR67_R1.contigs.groups BR67.contigs.groups)

mothur > summary.seqs(fasta=BR67.trim.contigs.fasta)

mothur > screen.seqs(fasta=BR67.trim.contigs.fasta, group=BR67.contigs.groups,
maxambig=0, maxlength=275, processors=32)

mothur > unique.seqs(fasta=BR67.trim.contigs.good.fasta)

mothur > count.seqs(name=BR67.trim.contigs.good.names,
group=BR67.contigs.good.groups)

mothur > summary.seqs(count=BR67.trim.contigs.good.count_table)

align.seqs(fasta=BR67.trim.contigs.good.unique.fasta, reference=silva.V34.fasta)

summary.seqs(fasta=BR67.trim.contigs.good.unique.align,
count=BR67.trim.contigs.good.count_table)

screen.seqs(fasta=BR67.trim.contigs.good.unique.align,
count=BR67.trim.contigs.good.count_table,
summary=BR67.trim.contigs.good.unique.summary, start=7435, end=17012,
maxhomop=8)

filter.seqs(fasta=BR67.trim.contigs.good.unique.good.align, vertical=T, trump=.)

unique.seqs(fasta=BR67.trim.contigs.good.unique.good.filter.fasta,
count=BR67.trim.contigs.good.good.count_table)

pre.cluster(fasta=BR67.trim.contigs.good.unique.good.filter.unique.fasta,
count=BR67.trim.contigs.good.unique.good.filter.count_table, diffs=2)

chimera.uchime(fasta=BR67.trim.contigs.good.unique.good.filter.unique.precluster.fasta,
count=BR67.trim.contigs.good.unique.good.filter.unique.precluster.count_table,
dereplicate=t)

remove.seqs(fasta=BR67.trim.contigs.good.unique.good.filter.unique.precluster.fasta,
accnos=BR67.trim.contigs.good.unique.good.filter.unique.precluster.denovo.uchime.accnos,
count=BR67.trim.contigs.good.unique.good.filter.unique.precluster.denovo.uchime.pick.count_table)

summary.seqs(fasta=current, count=current)

classify.seqs(fasta=BR67.trim.contigs.good.unique.good.filter.unique.precluster.pick.fasta,
count=BR67.trim.contigs.good.unique.good.filter.unique.precluster.denovo.uchime.pick.count_table, reference=trainset16_022016.pds.fasta,
taxonomy=trainset16_022016.pds.tax, cutoff=80)

remove.lineage(fasta=BR67.trim.contigs.good.unique.good.filter.unique.precluster.pick.fasta,
count=BR67.trim.contigs.good.unique.good.filter.unique.precluster.denovo.uchime.pick.count_table,
taxonomy=BR67.trim.contigs.good.unique.good.filter.unique.precluster.pick.pds.wang.taxonomy, taxon=Chloroplast-Mitochondria-unknown-Archaea-Eukaryota)

system(mv BR67.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.fasta BR67.final.fasta)

system(mv BR67.trim.contigs.good.unique.good.filter.unique.precluster.pick.pds.wang.pick.taxonomy BR67.final.rdp.taxonomy)

system(mv BR67.trim.contigs.good.unique.good.filter.unique.precluster.denovo.uchime.pick.pick.count_table BR67.final.count_table)

dist.seqs(fasta=BR67.final.fasta, cutoff=0.15, processors=30)

cluster(column=BR67.final.dist, count=BR67.final.count_table)

make.shared(list=BR67.final.opti_mcc.list, count=BR67.final.count_table, label=0.01)

remove.rare(shared=BR67.final.opti_mcc.shared, count=BR67.final.count_table, list=BR67.final.opti_mcc.list, nseqs=10)

list.seqs(list=BR67.final.opti_mcc.0.03.pick.list)

classify.otu(list=BR67.final.opti_mcc.0.03.pick.list, count=BR67.final.count_table, taxonomy=BR67.final.rdp.taxonomy, label=0.01)

get.seqs(accnos=BR67.final.opti_mcc.0.03.pick.accnos, taxonomy=BR67.final.rdp.taxonomy)

phylotype(taxonomy=BR67.final.rdp.pick.taxonomy, cutoff=6)

make.shared(list=BR67.final.rdp.pick.tx.list, count=BR67.final.pick.count_table, label=1-3-4-5)

classify.otu(list=BR67.final.rdp.pick.tx.list, count=BR67.final.pick.count_table, taxonomy=BR67.final.rdp.pick.taxonomy, label=1-3-4-5)

count.groups(shared=BR67.final.opti_mcc.0.03.pick.shared)

sub.sample(shared=BR67.final.opti_mcc.0.03.pick.shared, size=42595)

make.biom(shared=BR67.final.opti_mcc.0.03.pick.0.03.subsample.shared, constaxonomy=BR67.final.opti_mcc.0.03.pick.0.03.cons.taxonomy, groups=BR67_BB_A06-BR67_BA_B03-BR67_D3_B05-BR67_D0_T8_A05-BR67_D3_B04-BR67_D0_T4_B01-BR67_D0_T4_A06-BR67_BA_B06-BR67_D0_T8_B03-BR67_D0_T4_A01-BR67_BB_A04-BR67_BA_A06-BR67_BA_A04-BR67_BB_A01-BR67_D0_T4_A02-BR67_BB_B02-BR67_BB_B01-BR67_D0_T8_A04-BR67_BA_B04-BR67_D0_T8_A06-BR67_D0_T4_B02-BR67_BA_A05-BR67_BB_B03-BR67_BB_B04-BR67_BA_B05-BR67_BA_A01-BR67_BB_B06-BR67_BA_A02-BR67_BB_A02-BR67_D0_T8_A03-BR67_D0_T4_B04-BR67_BA_A03-BR67_D0_T4_B03-BR67_D0_T4_B06-BR67_BA_B01-BR67_D0_T4_A04-BR67_D0_T8_A02-BR67_D3_B03-BR67_BB_B05-BR67_D0_T8_A01-BR67_D0_T4_A05-BR67_D0_T4_B05-BR67_BB_A03-BR67_D3_B06-BR67_D0_T4_A03-BR67_D0_T8_B02-BR67_BA_B02-BR67_D1_B01-BR67_D1_B05-BR67_D3_A02-BR67_D1_A01-BR67_BB_A05-BR67_D3_B01-BR67_D1_A04-BR67_D0_T8_B05-BR67_D1_A06-BR67_D3_A06-BR67_D0_T8_B04-BR67_D3_A05-BR67_D1_B03-BR67_D1_B02-BR67_D0_T8_B01-BR67_D1_A03-BR67_D1_B04-BR67_D3_B02-BR67_D1_B06-BR67_D1_A02-BR67_D1_A05-BR67_D3_A03-BR67_D3_A01-BR67_D2_A04-

BR67_D3_A04-BR67_D2_B03-BR67_D2_B02-BR67_D0_T8_B06-BR67_D2_A05-BR67_D2_B01-BR67_D2_B05-BR67_D2_A03-BR67_D2_A02-BR67_D2_B04-BR67_D2_A06-BR67_D2_A01)

rarefaction.single(shared=BR67.final.opti_mcc.0.03.pick.0.03.subsample.shared, freq=500, processors=32)

rarefaction.single(shared=BR67.final.opti_mcc.0.03.pick.shared, freq=500, processors=32)

summary.single(shared=BR67.final.opti_mcc.0.03.pick.0.03.subsample.shared, label=0.03, calc=coverage-nseqs)

summary.single(shared=BR67.final.opti_mcc.0.03.pick.0.03.subsample.shared, calc=nseqs-coverage-sobs-invsimpson)

dist.shared(shared=BR67.final.opti_mcc.0.03.pick.0.03.subsample.shared, calc=braycurtis, subsample=42595, processors=32, groups=BR67_BB_A06-BR67_BA_B03-BR67_D3_B05-BR67_D0_T8_A05-BR67_D3_B04-BR67_D0_T4_B01-BR67_D0_T4_A06-BR67_BA_B06-BR67_D0_T8_B03-BR67_D0_T4_A01-BR67_BB_A04-BR67_BA_A06-BR67_BA_A04-BR67_BB_A01-BR67_D0_T4_A02-BR67_BB_B02-BR67_BB_B01-BR67_D0_T8_A04-BR67_BA_B04-BR67_D0_T8_A06-BR67_D0_T4_B02-BR67_BA_A05-BR67_BB_B03-BR67_BB_B04-BR67_BA_B05-BR67_BA_A01-BR67_BB_B06-BR67_BA_A02-BR67_BB_A02-BR67_D0_T8_A03-BR67_D0_T4_B04-BR67_BA_A03-BR67_D0_T4_B03-BR67_D0_T4_B06-BR67_BA_B01-BR67_D0_T4_A04-BR67_D0_T8_A02-BR67_D3_B03-BR67_BB_B05-

BR67_D0_T8_A01-BR67_D0_T4_A05-BR67_D0_T4_B05-BR67_BB_A03-BR67_D3_B06-BR67_D0_T4_A03-BR67_D0_T8_B02-BR67_BA_B02-BR67_D1_B01-BR67_D1_B05-BR67_D3_A02-BR67_D1_A01-BR67_BB_A05-BR67_D3_B01-BR67_D1_A04-BR67_D0_T8_B05-BR67_D1_A06-BR67_D3_A06-BR67_D0_T8_B04-BR67_D3_A05-BR67_D1_B03-BR67_D1_B02-BR67_D0_T8_B01-BR67_D1_A03-BR67_D1_B04-BR67_D3_B02-BR67_D1_B06-BR67_D1_A02-BR67_D1_A05-BR67_D3_A03-BR67_D3_A01-BR67_D2_A04-BR67_D3_A04-BR67_D2_B03-BR67_D2_B02-BR67_D0_T8_B06-BR67_D2_A05-BR67_D2_B01-BR67_D2_B05-BR67_D2_A03-BR67_D2_A02-BR67_D2_B04-BR67_D2_A06-BR67_D2_A01)

APPENDIX B

MULTIVARIATE ANALYSIS

Analysis of variance using distance matrices (ADONIS)

This uses psuedo F-ratios and permutation to determine significance of groupings based on centroids. In other words, significance denotes a difference in centroids for different groupings of sample communities.

Comparison	Index	Adonis Output	Beta Dispersion
Donor 1 : Donor 2	**Bray-Curtis**	Df SumsOfSqs MeanSqs F.Model R2 Pr(>F) Donor 1 8.0781 8.0781 43.132 0.35029 0.000999 *** Residuals 80 14.9831 0.1873 0.64971 Total 81 23.0611 1.00000	0.01499 *
	Jaccard	Df SumsOfSqs MeanSqs F.Model R2 Pr(>F) Donor 1 6.9202 6.9202 25.995 0.24525 0.000999 *** Residuals 80 21.2971 0.2662 0.75475 Total 81 28.2173 1.00000	0.008991 **
	Weighted UniFrac	Df SumsOfSqs MeanSqs F.Model R2 Pr(>F) Donor 1 2.6069 2.60693 44.902 0.3595 0.000999 *** Residuals 80 4.6447 0.05806 0.6405 Total 81 7.2516 1.0000	0.1259
	UnWeigted UniFrac	Df SumsOfSqs MeanSqs F.Model R2 Pr(>F) Donor 1 4.1885 4.1885 17.985 0.18355 0.000999 *** Residuals 80 18.6310 0.2329 0.81645 Total 81 22.8195 1.00000	0.02098 *
Donor1 : Donor 2 No Outlier	**Bray-Curtis**	Df SumsOfSqs MeanSqs F.Model R2 Pr(>F) Donor 1 8.093 8.093 50.904 0.41083 0.000999 *** Residuals 73 11.606 0.159 0.58917 Total 74 19.699 1.00000	0.5115
	Jaccard	Df SumsOfSqs MeanSqs F.Model R2 Pr(>F) Donor 1 6.9321 6.9321 28.544 0.2811 0.000999 *** Residuals 73 17.7283 0.2429 0.7189 Total 74 24.6604 1.0000	0.3966

Table 5: Adonis and Beta-disperion calculations for various data groupings. Significance p-values are reported.

Table 5 Continued

Donor1 : Donor 2 No Outlier	**Weighted UniFrac**	Df SumsOfSqs MeanSqs F.Model R2 Pr(>F) Donor 1 2.5369 2.53691 53.007 0.42067 0.000999 *** Residuals 73 3.4938 0.04786 0.57933 Total 74 6.0307 1.00000	0.6813
	UnWeighted UniFrac	Df SumsOfSqs MeanSqs F.Model R2 Pr(>F) Donor 1 4.1126 4.1126 18.926 0.20588 0.000999 *** Residuals 73 15.8632 0.2173 0.79412 Total 74 19.9758 1.00000	0.6064
Donor 1 by Tgroups + Reactor	**Bray-Curtis**	Df SumsOfSqs MeanSqs F.Model R2 Pr(>F) Tgroup 1 1.5351 1.53508 33.573 0.17245 0.000999 *** Reactor 4 5.7661 1.44152 31.527 0.64777 0.000999 *** Residuals 35 1.6003 0.04572 0.17978 Total 40 8.9015 1.00000	0.000999 *** 0.2118
	Jaccard	Df SumsOfSqs MeanSqs F.Model R2 Pr(>F) Tgroup 1 1.7155 1.71549 17.365 0.14414 0.000999 *** Reactor 4 6.7285 1.68213 17.027 0.56534 0.000999 *** Residuals 35 3.4577 0.09879 0.29052 Total 40 11.9017 1.00000	0.000999 *** 0.2408
	Weighted UniFrac	Df SumsOfSqs MeanSqs F.Model R2 Pr(>F) Tgroup 1 0.54468 0.54468 32.316 0.17956 0.000999 *** Reactor 4 1.89883 0.47471 28.164 0.62597 0.000999 *** Residuals 35 0.58992 0.01685 0.19447 Total 40 3.03343 1.00000	0.000999 *** 0.1938
	UnWeighted UniFrac	Df SumsOfSqs MeanSqs F.Model R2 Pr(>F) Tgroup 1 1.5189 1.51895 14.358 0.15363 0.000999 *** Reactor 4 4.6652 1.16631 11.024 0.47186 0.000999 *** Residuals 35 3.7028 0.10579 0.37451 Total 40 9.8870 1.00000	0.000999 *** 0.000999 ***
Donor 1 Autologous by Reactor + Timepoint	**Bray-Curtis**	Df SumsOfSqs MeanSqs F.Model R2 Pr(>F) Reactor 2 1.89951 0.94976 33.83 0.68068 0.000999 *** as.factor(Timepoint) 6 0.55420 0.09237 3.29 0.19859 0.002997 ** Residuals 12 0.33689 0.02807 0.12072 Total 20 2.79060 1.00000	0.1808 0.9241
	Jaccard	Df SumsOfSqs MeanSqs F.Model R2 Pr(>F) Reactor 2 2.5155 1.25774 14.8919 0.57056 0.000999 *** as.factor(Timepoint) 6 0.8798 0.14664 1.7362 0.19956 0.037962 * Residuals 12 1.0135 0.08446 0.22988 Total 20 4.4088 1.00000	0.1918 0.9391

Table 5 Continued

Donor 1 Autologous by Reactor + Timepoint	**Weighted UniFrac**	Df SumsOfSqs MeanSqs F.Model R2 Pr(>F) Reactor 2 0.59716 0.298580 32.570 0.64759 0.000999 *** as.factor(Timepoint) 6 0.21496 0.035827 3.908 0.23312 0.000999 *** Residuals 12 0.11001 0.009167 0.11930 Total 20 0.92213 1.00000 ---	0.1419 0.9101
	UnWeighted UdniFrac	Df SumsOfSqs MeanSqs F.Model R2 Pr(>F) Reactor 2 1.4054 0.70269 6.2077 0.37298 0.000999 *** as.factor(Timepoint) 6 1.0042 0.16737 1.4786 0.26652 0.018981 * Residuals 12 1.3584 0.11320 0.36050 Total 20 3.7680 1.00000	0.02398 * 0.9251
Donor 1 NonAutologous by Reactor + Timepoint	**Bray-Curtis**	Df SumsOfSqs MeanSqs F.Model R2 Pr(>F) Reactor 2 3.8666 1.93328 57.347 0.84500 0.000999 *** as.factor(Timepoint) 6 0.3384 0.05640 1.673 0.07395 0.126873 Residuals 11 0.3708 0.03371 0.08104 Total 19 4.5758 1.00000	0.2807 0.999
	Jaccard	Df SumsOfSqs MeanSqs F.Model R2 Pr(>F) Reactor 2 4.2130 2.10651 24.8572 0.72923 0.000999 *** as.factor(Timepoint) 6 0.6322 0.10536 1.2433 0.10942 0.256743 Residuals 11 0.9322 0.08474 0.16135 Total 19 5.7774 1.00000	0.2637 0.996
	Weighted UniFrac	Df SumsOfSqs MeanSqs F.Model R2 Pr(>F) Reactor 2 1.31443 0.65721 59.241 0.84751 0.000999 *** as.factor(Timepoint) 6 0.11446 0.01908 1.720 0.07380 0.116883 Residuals 11 0.12203 0.01109 0.07868 Total 19 1.55092 1.00000	0.1928 0.997
	UnWeightd UniFrac	Df SumsOfSqs MeanSqs F.Model R2 Pr(>F) Reactor 2 3.2282 1.61410 22.083 0.70954 0.000999 *** as.factor(Timepoint) 6 0.5175 0.08625 1.180 0.11374 0.303696 Residuals 11 0.8040 0.07309 0.17672 Total 19 4.5497 1.00000	0.000999 *** 0.966
Donor 2 by Tgroups + Reactor	**Bray-Curtis**	Df SumsOfSqs MeanSqs F.Model R2 Pr(>F) Tgroup 1 1.5351 1.53508 33.573 0.17245 0.000999 *** Reactor 4 5.7661 1.44152 31.527 0.64777 0.000999 *** Residuals 35 1.6003 0.04572 0.17978 Total 40 8.9015 1.00000	0.000999 *** 0.2418
	Jaccard	Df SumsOfSqs MeanSqs F.Model R2 Pr(>F) Tgroup 1 1.7155 1.71549 17.365 0.14414 0.000999 *** Reactor 4 6.7285 1.68213 17.027 0.56534 0.000999 *** Residuals 35 3.4577 0.09879 0.29052 Total 40 11.9017 1.00000	0.001998 ** 0.2448

Table 5 Continued

Donor 2 by Tgroups + Reactor	**Weighted UniFrac**	Df SumsOfSqs MeanSqs F.Model R2 Pr(>F) Tgroup 1 0.53327 0.53327 34.868 0.19445 0.000999 *** Reactor 4 1.67389 0.41847 27.362 0.61036 0.000999 *** Residuals 35 0.53530 0.01529 0.19519 Total 40 2.74246 1.00000	0.000999 *** 0.0989 .
	Weighted UniFrac	Df SumsOfSqs MeanSqs F.Model R2 Pr(>F) Tgroup 1 1.5190 1.51899 14.387 0.15374 0.000999 *** Reactor 4 4.6656 1.16640 11.047 0.47223 0.000999 *** Residuals 35 3.6954 0.10558 0.37403 Total 40 9.8800 1.00000	0.000999 *** 0.000999 ***
Donor 2 Autologous by Reactor + Timepoint	**Bray-Curtis**	Df SumsOfSqs MeanSqs F.Model R2 Pr(>F) Reactor 2 0.82472 0.41236 20.5268 0.63077 0.000999 *** as.factor(Timepoint) 6 0.26179 0.04363 2.1719 0.20022 0.022977 * Residuals 11 0.22098 0.02009 0.16901 Total 19 1.30748 1.00000	0.2318 0.7003
	Jaccard	Df SumsOfSqs MeanSqs F.Model R2 Pr(>F) Reactor 2 1.43354 0.71677 12.6684 0.55589 0.000999 *** as.factor(Timepoint) 6 0.52288 0.08715 1.5403 0.20276 0.098901 . Residuals 11 0.62238 0.05658 0.24134 Total 19 2.57880 1.00000	0.2428 0.5664
	Weighted UniFrac	Df SumsOfSqs MeanSqs F.Model R2 Pr(>F) Reactor 2 0.150793 0.075396 18.8072 0.60747 0.000999 *** as.factor(Timepoint) 6 0.053339 0.008890 2.2175 0.21488 0.038961 * Residuals 11 0.044098 0.004009 0.17765 Total 19 0.248230 1.00000	0.1918 0.8621
	UnWeighted UniFrac	Df SumsOfSqs MeanSqs F.Model R2 Pr(>F) Reactor 2 1.9035 0.95176 9.8221 0.50689 0.000999 *** as.factor(Timepoint) 6 0.7859 0.13098 1.3517 0.20927 0.093906 . Residuals 11 1.0659 0.09690 0.28384 Total 19 3.7553 1.00000	0.5694 0.1828
Donor 2 NonAutologous by Reactor + Timepoint	**Bray-Curtis**	Df SumsOfSqs MeanSqs F.Model R2 Pr(>F) Reactor 2 2.6298 1.31488 61.079 0.81584 0.000999 *** as.factor(Timepoint) 6 0.3353 0.05588 2.596 0.10402 0.007992 ** Residuals 12 0.2583 0.02153 0.08014 Total 20 3.2234 1.00000	0.006993 ** 0.9491
	Jaccard	Df SumsOfSqs MeanSqs F.Model R2 Pr(>F) Reactor 2 3.4133 1.70663 27.4352 0.72038 0.000999 *** as.factor(Timepoint) 6 0.5784 0.09640 1.5497 0.12208 0.131868 Residuals 12 0.7465 0.06221 0.15755 Total 20 4.7381 1.00000	0.000999 *** 0.956
	Weighted UniFrac	Df SumsOfSqs MeanSqs F.Model R2 Pr(>F) Reactor 2 1.02439 0.51219 77.612 0.82621 0.000999 *** as.factor(Timepoint) 6 0.13629 0.02271 3.442 0.10992 0.003996 ** Residuals 12 0.07919 0.00660 0.06387 Total 20 1.23987 1.00000	0.01598 * 0.983

Table 5 Continued

Donor 2 NonAutologous by Reactor + Timepoint	**UnWeighted UniFrac**	Df SumsOfSqs MeanSqs F.Model R2 Pr(>F) Reactor 2 2.0137 1.00687 12.9436 0.56857 0.000999 *** as.factor(Timepoint) 6 0.5946 0.09909 1.2739 0.16787 0.176823 Residuals 12 0.9335 0.07779 0.26356 Total 20 3.5418 1.00000	0.2158 0.998

Table 5: Adonis and Beta-disperion calculations for various data groupings. Significance p-values are reported.

Adonis Pairwise Comparisons

Subset	Comparison	Adonis p-value (Bray-Curtis based)	B-disperion p-value	Adjusted Adonis p-value (bonferroni)	Adjusted B-dispersion p-value (TukeyHSD)
Donor 1	Self1 : Self2	0.000999 ***	0.4456	0.014985	0.9682403
	Self1 : Self3	0.003996 **	0.0979 .	0.05994	0.4458814
	Self1 : NonSelf1	0.000999 ***	0.04196 *	0.014985	0.2772458
	Self1 : NonSelf2	0.000999 ***	0.2887	0.014985	0.748485
	Self1 : NonSelf3	0.001998 **	0.7772	0.02997	0.9991732
	Self2 : Self3	0.002997 **	0.2048	0.044955	0.8901624
	Self2 : NonSelf1	0.000999 ***	0.04296 *	0.014985	0.7357925
	Self2 : NonSelf2	0.000999 ***	0.4995	0.014985	0.9897587
	Self2 : NonSelf3	0.002997 **	0.7063	0.044955	0.9979994
	Self3 : NonSelf1	0.000999 ***	0.6164	0.014985	0.9995465
	Self3 : NonSelf2	0.001998 **	0.6873	0.02997	0.9981959
	Self3 : NonSelf3	0.000999 ***	0.2158	0.014985	0.667828

Table 6 Continued

	NonSelf1 : NonSelf2	0.002997 **	0.4545	0.044955	0.9778798
	NonSelf1 : NonSelf3	0.001998 **	0.08492 .	0.02997	0.4708691
	NonSelf2 : NonSelf3	0.001998 **	0.4525	0.02997	0.9081874
Donor 2	Self1 : Self2	0.000999 ***	0.07093 .	0.014985	0.3266871
	Self1 : Self3	0.002997 **	0.3876	0.044955	0.9551765
	Self1 : NonSelf1	0.000999 ***	0.008991 **	0.014985	0.4560982
	Self1 : NonSelf2	0.000999 ***	0.08691 .	0.014985	0.7941514
	Self1 : NonSelf3	0.001998 **	0.1598	0.02997	0.8137610
	Self2 : Self3	0.000999 ***	0.4296	0.014985	0.8633881
	Self2 : NonSelf1	0.000999 ***	0.8831	0.014985	0.9999015
	Self2 : NonSelf2	0.000999 ***	0.5295	0.014985	0.9675412
	Self2 : NonSelf3	0.001998 **	0.01998 *	0.02997	0.0244037
	Self3 : NonSelf1	0.000999 ***	0.3067	0.014985	0.9405103
	Self3 : NonSelf2	0.000999 ***	0.8142	0.014985	0.9989229
	Self3 : NonSelf3	0.000999 ***	0.1159	0.014985	0.3390353
	NonSelf1 : NonSelf2	0.001998 **	0.3267	0.02997	0.9930625
	NonSelf1 : NonSelf3	0.001998 **	0.001998 **	0.02997	0.0428820
	NonSelf2 : NonSelf3	0.001998 **	0.01099 *	0.02997	0.1458756

Table 6: Post hoc pairwise adonis and beta-dispersion analysis. Results with adjusted p-values.

APPENDIX C

VIRAL FILTRATE METAGENOMICS

Input VLP DNA Isolation and Sequencing

DNA was extracted from 400ul of viral fractions using a Purelink Viral DNA/RNA miniprep kit with an extended proteinase K step of 2 hours. DNA integrity was visualized with agarose gel electrophoresis and DNA concentration was quantified with using a Qubit fluorimeter high sensitivity DNA assay. Metagenomic sequencing was performed on an Illumina MiSeq platform generating 300bp paired end reads with V3 chemistry to a total of 25 million reads across six samples (4 MBRA and 2 fecal samples).

VLP DNA Extraction for MBRA Samples 01/05/2017

Same purelink viral DNA protocol used 11/22/2016.

Testing: 2 x (55X) concentrates AO5+AO6 & BO5+BO6

Concentrates – 90uL concentrate into 90uL SM Buffer for 180 uL

Set a heatblock to 56 C

1. DNAse treat samples(RQ1 Promega)
    a. Concentrates (20uL RQ1)

    Incubate at 37 C for 30 minutes

    Inactivate: Add 5ul of .5M EDTA and vortex. Spin down (*use small spinner not centrifuge, this is only to bring droplets off of cap and sides after vortexing)

    Incubate at 65 C for 10 minutes. Spin down (*same as above)

2. Proteinase K/Lysis Buffer
    a. Concentrates (25uL Proteinase K, 200uL Lysis Buffer w/ 5.6ug Carrier RNA)

Vortex 15 seconds and spin down(*) then incubate at 56 C for 2 hours

Add 96-100% EtOH and vortex for 15 seconds

a. Concentrates (250uL)

Incubate lysate for 5 minutes at room temperature

3. Add lysate to spin column, centrifuge at 6800 x g for 1 minute. Discard collection tube and place column in wash tube
4. Wash with 500uL Wash Buffer(WII) w/ EtOH. Centrifuge at 6800 x g for 1 minute and discard flow through.
5. Repeat step 3. Discard wash tube and place column in new wash tube.
6. Centrifuge @ maximum speed for 1 minute to remove residual wash buffer.
7. Place spin column in a clean 1.7 mL recovery tube.
8. Elute with 50 uL TE. Add TE to center of column and let stand for 1 minute. Centrifuge at maximum speed for 1 minute. Discard spin column and store DNA
9. Qubit to quantify

RESULTS

Qubit dsHS assay using 5uL of sample

Prep	Qubit assay concentration ng/ml	Actual concentration ng/ml
AO5 + AO6	18	720
AO5 + AO6	16.9	676
AO5 + AO6	16.8	672
BO5 + BO6	43.9	1760
BO5 + BO6	43.4	1740
BO5 + BO6	43.1	1720

Table 7: MBRA.1 Viral DNA [] for viral metagenomics. Samples were quantified using a Qubit dsHS DNA assay and 5ul of sample.

*Notes from preparation: New concentration prep seems to have provided enough material for a good extraction.

VLP DNA Extraction for MBRA Samples 01/06/2017

Same purelink viral DNA protocol used 11/22/2016.

Testing: 2 x (55X) concentrates AO2+AO3 & BO2+BO3

Set a heatblock to 56 C

1. DNAse treat samples(RQ1 Promega)
    a. Concentrates (20uL RQ1)

    Incubate at 37 C for 30 minutes

    Inactivate: Add 5ul of .5M EDTA and vortex. Spin down (*use small spinner not centrifuge, this is only to bring droplets off of cap and sides after vortexing)

    Incubate at 65 C for 10 minutes. Spin down (*same as above)

2. Proteinase K/Lysis Buffer
    a. Concentrates (25uL Proteinase K, 200uL Lysis Buffer w/ 5.6ug Carrier RNA)

    Vortex 15 seconds and spin down(*) then incubate at 56 C for 2 hours

    Add 96-100% EtOH and vortex for 15 seconds
    a. Concentrates (250uL)

    Incubate lysate for 5 minutes at room temperature
3. Add lysate to spin column, centrifuge at 6800 x g for 1 minute. Discard collection tube and place column in wash tube
4. Wash with 500uL Wash Buffer(WII) w/ EtOH. Centrifuge at 6800 x g for 1 minute and discard flow through.
5. Repeat step 3. Discard wash tube and place column in new wash tube.
6. Centrifuge @ maximum speed for 1 minute to remove residual wash buffer.
7. Place spin column in a clean 1.7 mL recovery tube.
8. Elute with 50 uL TE. Add TE to center of column and let stand for 1 minute. Centrifuge at maximum speed for 1 minute. Discard spin column and store DNA @
9. Qubit to quantify

RESULTS

Qubit dsHS assay using 5uL of sample

Prep	Qubit assay concentration ng/ml	Actual concentration ng/ml
AO2 + AO3	28.7	1150
AO2 + AO3	29.2	1170
AO2 + AO3	29.3	1170
BO2 + BO3	11.6	464
BO2 + BO3	11.8	472
BO2 + BO3	11.8	472
SM Buffer (Used in all preps since 11/22/16)	Too Low to Detect (<.5ng/ml)	
TE Buffer (Used in all preps since 11/22/16)	Too Low to Detect (<.5ng/ml)	

Table 8: MBRA.2 Viral DNA [] for viral metagenomics. Samples were quantified using a Qubit dsHS DNA assay and 5ul of sample.

*Notes from preparation: New concentration prep seems to have provided enough material for a good extraction.

Concentrates – 90uL concentrate into 90uL SM Buffer for 180 uL

Imaging 01/12/2017

This should have been done directly after extraction next time. Run 10ul on 1% gel and image with several millisecond exposure to capture faint bands. Lane 5 may have some degradation which appears as very light smearing. This may affect sequencing. 10ul of ladder was loaded which accounts for extra brightness in lane 1.

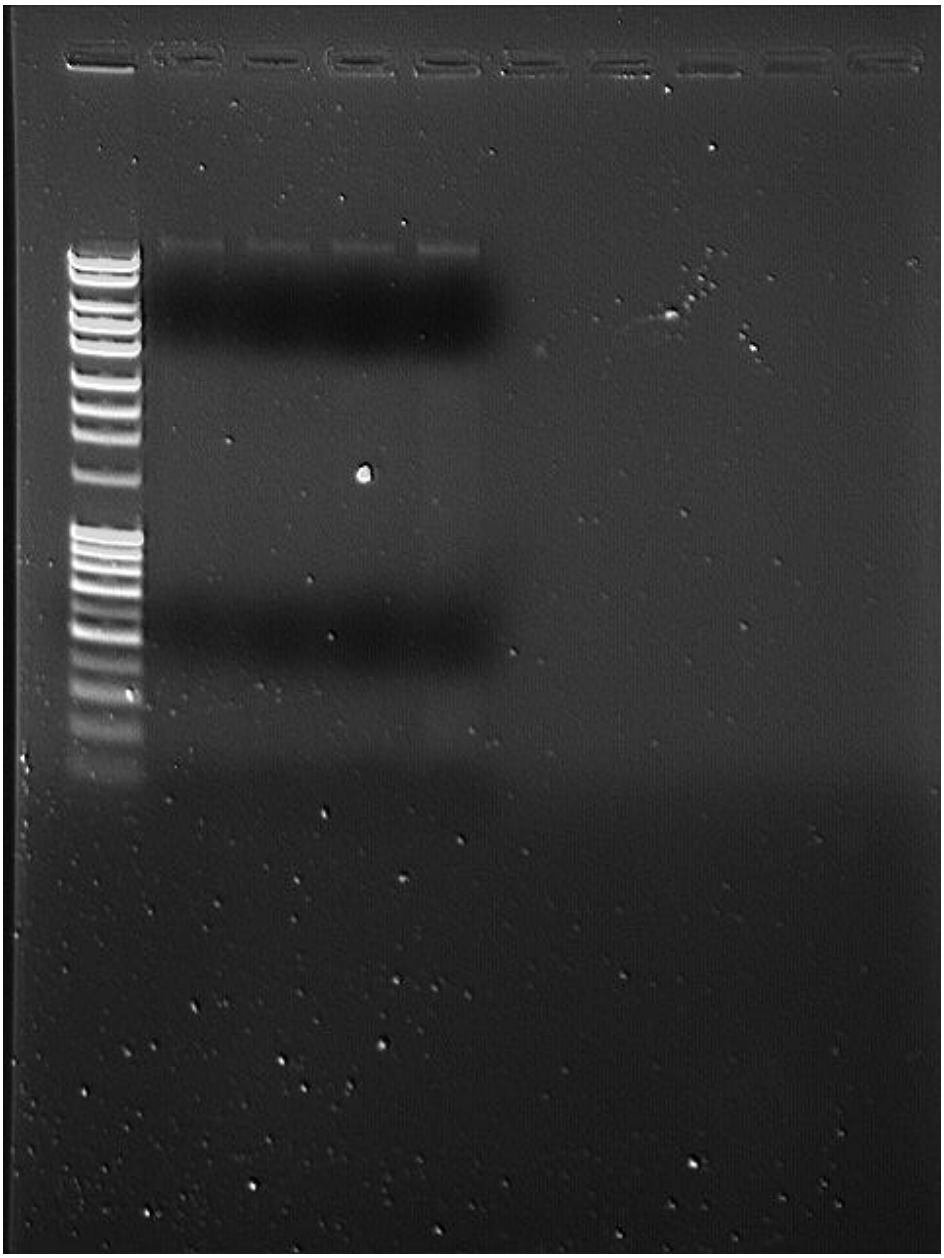

Figure 15: Viral DNA integrity gel. Lanes 1- (1kb + Ladder), 2-(A02+A03), 3-(B02+B03), 4- (A05+A06), 5- (B05+B06).

Since we are waiting on other sequencing samples for a few more weeks these need to be frozen and not thawed until they arrive at the sequencing center. Each will be aliquoted into 10ul aliquots in sterile tubes in the biosafety cabinet and frozen.

VLP Sequence Analysis

Sequences were quality trimmed via Trimmomatic version 0.36 with sliding window parameters of 4:15, minimum length 50 and average quality greater than 30 then filtered for lengths greater than 250bp. All high quality reads between 250 and 300bp's were mapped against an existing human GI phageome database using Bowtie2 with high stringency, and end-to-end parameters. This contig database was generated through longitudinal study of the phage fractions of fecal samples from the two donors. Given results of previous quality tests, single read mappings with these parameters were treated as positive hits (46). Raw reads were queried against the Silva 16S rRNA database with BLASTn.

De novo assembly was performed by modifying IDBA-UD assembler to maximum k-mer lengths of 300bp and augmented read lengths of the same. Additionally a cross-assembly approach was adopted in the hopes of capturing low abundance contigs. A minimum cutoff of 1000 bp was applied to all contigs as has been adopted in other metagenomic virome studies. Raw quality trimmed reads from each sample were then mapped to the co-assembly contig database. Open reading frames were extracted from contigs using WebMGA (88). These were queried against the most recent Phage Orthologous Group database via psiBLAST to determine the existence of phage specific proteins.

Community analysis was accomplished by grouping contigs based on sequence similarity using the Heterologous Viral Diversity Index procedure developed by Dr. David Pride's research group (81).

APPENDIX D

PCR PRIMERS & TARGETS

		Start	Stop	Length	Tm	GC (%)	Amplicon Length
nc_156.1 Fw	GGGAAGGAACTGCACCTTATAG	15889	15911	22	62	50	171
nc_156.1 Rev	TCCTTCTGCCTCAAACTCATC	16039	16060	21	62	47.6	
nc_156.2 Fw	CGCATAGGGCAGAGAAGTAAA	25505	25526	21	62	47.6	194
nc_156.2 Rev	AGGTGGAAATGAGCGATGAG	25679	25699	20	62	50	
nc_419.1 Rev	GGGTTCGTCTTTTGGGCAAC	5894	5875	20	60	55	191
nc_419.1 Fw	TTGCAAGCATGGCGGAAAAA	5704	5723	20	59.9	45	
nc_63.1 Fw	TCTGTTTGCGAGCAACTTGC	14730	14749	20	60	50	158
nc_63.1 Rev	GCCGAAAGAATCACGATGCC	14887	14868	20	60	55	
mb_61.1 Fw	CTCGCTGACAATGATGCACG	63615	63634	20	60	55	193
mb_61.1 Rev	GTAGACTGAACGGGTGGCAA	63807	63788	20	60	55	
mb_924.1 Fw	TGTAGCCGGCACATTTGACT	15127	15146	20	60	50	175
mb_924.1 Rev	GCTTGTGCTGGACTGGTACT	15301	15282	20	60	55	

Table 9: List of generated primers.

Nc25 (all_hybrid_55) Tm 59°C

Nc51 (all_hybrid_39) Tm 68 °C

Primer	Sequence
ALL_nc_81_4103F	GTTTCACCCCTTTCTTGTTG
ALL_nc_81_4413R	AGTATCTATCCAGCAAGGGT
ALL_hybrid_nc_375_5796F	CTCATACATTGCACTCATCCTC
ALL_hybrid_nc_375_196R	CAGAAGAAATATTTTGAACTGGCTG
ALL_hybrid_nc_375_460R	CTCAAAAGTGATAGAGCATATTACAGAC
ALL_hybrid_nc_74_6766F	GAAATGGCTGGCTTTAAAACC
ALL_hybrid_nc_74_7271R	CTGAAGCAATCGGAGTTTTCC
ALL_nc_58_24634F	GAAATATCGTCCGGACAAGA
ALL_nc_58_24901R	AGGCTCCATCTCTGTATTCT
ALL_hybrid_55_65,950F	AGCCAGAACCAATTAAGAGG
ALL_hybrid_55_66,169R	TTGGTGGTAGTCTTTTCTCG

Table 10: List of Primers from HGP project.

Contig 81: Variants by mapping MSU timepoint 4 (MBRA spike) to HGP contig 81 (Originally assembled from first two MSU donor timepoints).

MSU1:

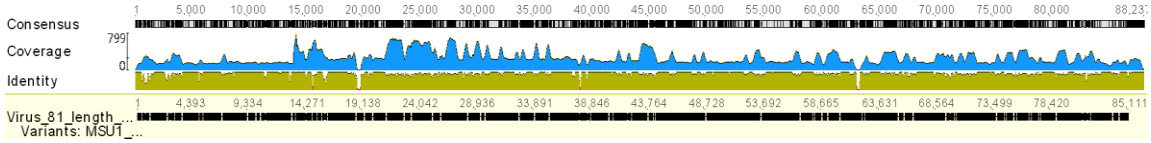

Figure 16: MSU1-Virus 81 coverage. Virus_81 varient output visualized in Geneious v 10.2.

78,708 reads mapped of 1,616,814 total reads

8842 Variations (SNPs and Indels) + 2 Large deletions (263bp from 19676 to 19938 bp on template strand)(232bp from 63,206 to 63,437 bp on template strand)

Variation	Occurances
Deletion	88
Deletion (Tandem Repeat)	15
Insertion	110
Insertion (Tandem Repeat)	17
SNP	154
SNP (Transition)	2987
SNP (Transversion)	3745
Substitution	1726

Table 11: MSU1-Virus 81 variations.

Length: 88,237

Sequences: 78,708

Identical Sites: 43,725 (49.6%)

Pairwise % Identity: 90.1%

Coverage of 88,237 bases:

Mean: 264.2 Std Dev: 130.8

Minimum: 1 Maximum: 799

Forward: ? Reverse: ?

Ref-Seq: 100% of 85,111 bp

MSU2:

310,971 reads mapped of 1,636,504 total reads

214 Variations (SNPs and Indels)

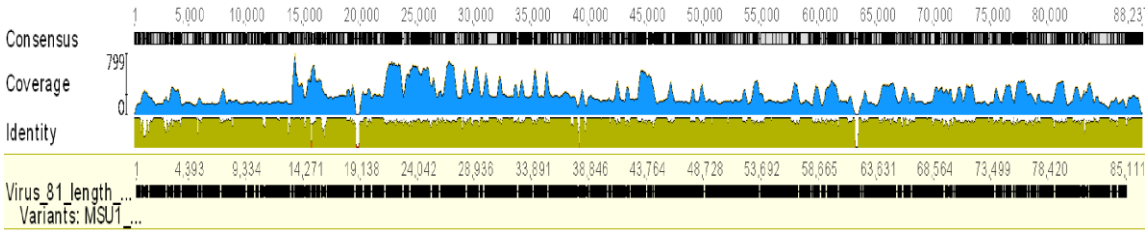

Figure 4: MSU2-Virus81 coverage. Virus_81 varient output visualized in Geneious v 10.2.

	Occurances
Deletion	1
Deletion (Tandem Repeat)	2
Insertion	4
Insertion (Tandem Repeat)	0
SNP	138
SNP (Transition)	48
SNP (Transversion)	11
Substitution	10

Table 12: MSU2-Virus 81 variations.

Length: 88,183

Sequences: 310,971

Identical Sites: 22,213 (25.3%)

Pairwise % Identity: 99.6%

Coverage of 88,183 bases:

Mean: 1034.7 Std Dev: 331.3

Minimum: 1 Maximum: 4279

Forward: ? Reverse: ?

Ref-Seq: 100% of 85,111 bp

Contig 25:

MSU 1:

97,588 reads mapped of 1,616,814 total reads

4189 variations (SNPs and Indels)

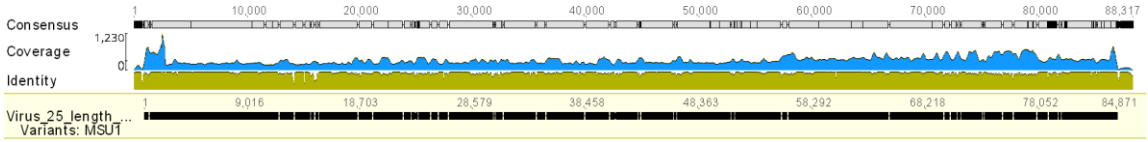

Figure 5: MSU1-Virus25 coverage. Virus_25 varient output visualized in Geneious v 10.2.

Variation	Occurrences
Deletion	20
Deletion(Tandem Repeat)	3
Insertion	52
Insertion(Tamdem Repeat)	9
SNP	173
SNP(Transition)	1696
SNP(Transversion)	1397
Substitution	839

Table 13: MSU1-Virus 25 variations.

Length: 88,317

Sequences: 97,588

Identical Sites: 40,909 (46.3%)

Pairwise % Identity: 93.3%

Coverage of 88,317 bases:

Mean: 324.7 Std Dev: 129.6

Minimum: 1 Maximum: 1230

Forward: ? Reverse: ?

Ref-Seq: 100% of 84,871 bp

MSU 2:

201,876 reads mapped of 1,636,504 total reads

13817 variations (SNPs and Indels). Multiple large deletion areas.

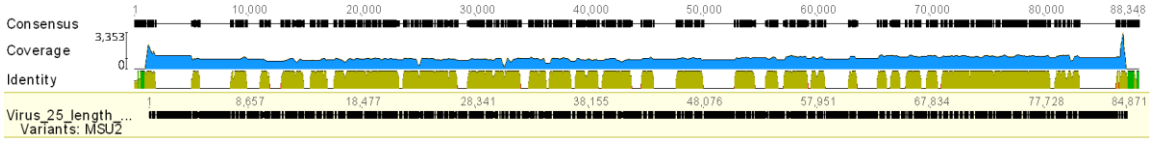

Figure 6: MSU2-Virus25 coverage. Virus_25 varient output visualized in Geneious v 10.2.

Variation	Occurrence
Deletion	249
Deletion(Tandem Repeat)	15
Insertion	234
Insertion(Tamdem Repeat)	16
SNP	43
SNP(Transition)	4333
SNP(Transversion)	5333
Substitution	3594

Table 14:MSU2-Virus 25 variations.

Length: 88,348

Sequences: 201,876

Identical Sites: 10,166 (11.6%)

Pairwise % Identity: 96.8%

Coverage of 88,348 bases:

Mean: 976.2 Std Dev: 252.8

Minimum: 1 Maximum: 3353

Forward: ? Reverse: ?

Ref-Seq: 100% of 84,871 bp

APPENDIX E

SANTIAGO-RODRIGUEZ ET AL. ASSEMBLY DATA

See reference Data Table:

Data table from Santiago-Rodriguez et al. available as supplemental material 2015.file:///C:/Users/b72t755/Documents/Pride_Chemostat_analysis/Pride_Chemostat_Supps.pdf.

Assembly	Contigs	n50	Max	Mean	Total Length	n80	Reads	Reads Aligned	
Don1_AllReactor	521	9421	60009	4598	2395739	2236	2509031	1066341	42.50011
Don2_AllReactor	527	5184	59067	3703	1951999	1962	2443339	1127260	46.13605
Don8_AllReactor	410	11386	76627	4431	1816854	2057	2312886	964757	41.71226
Don9_AllReactor	327	5963	71393	4033	1319115	2017	2380987	1053295	44.23775
DON10_ AllReactor	1677	4337	51354	2499	4191607	1740	2467899	843568	34.18163
Don1_Allreads	634	10741	70499	4934	3128440	2474	3737968	1688123	45.16152
Don2_Allreads	1125	6255	69476	4211	4737547	2222	3864218	1785169	46.19742
Don8_Allreads	511	10168	76630	4630	2366071	2255	3688775	1627434	44.11855
Don9_Allreads	484	7060	93790	4248	2056057	2202	3487768	1533455	43.96666
Don10_Allreads	1206	6718	62863	4467	5387235	2468	3805875	1552699	40.79743

Table 15: Modified IDBA-UD assembly statistics. Assembly statistics using IDBA-UD assembler with modified k-mer function to accommodate/iterate to the maximum read length.

Read mapping Pride viromes to…

*These were size filtered reads (min 250 bp, max 358) deduplicated.

454AllContigs NL10 (Ben's Yellowstone Viromes)

Commands run: bowtie2-build-s BowtieCombinedReferenceSequences.fasta myIndex

bowtie2-align-s -I 0 -X 800 -p 8 --sensitive -q -x myIndex -U forwardReads.fastq -S

bowtie.sam Bowtie command line duration: 56 seconds

234 of 4,229,034 reads were assembled to 454AllContigs to produce 44 contigs

4,228,800 reads were not assembled

MSU 4301 Contigs (Pilar's MSU viromes)

Commands run: bowtie2-build-s BowtieCombinedReferenceSequences.fasta myIndex bowtie2-align-s -I 0 -X 800 -p 8 --sensitive -q -x myIndex -U forwardReads.fastq -S bowtie.sam Bowtie command line duration: 92 seconds

213,068 of 4,229,034 reads were assembled to Gut_phageome_4301_file_PNAS_Manrique_et_al to produce 522 contigs 4,015,966 reads were not assembled