



Virome biogeography in the lower gastrointestinal tract of rhesus macaques with chronic diarrhea

Guoyan Zhao^{a,*}, Lindsay Droit^a, Margaret H. Gilbert^c, Faith R. Schiro^c, Peter J. Didier^c, Xuemei Si^d, Anne Paredes^a, Scott A. Handley^a, Herbert W. Virgin^{a,2}, Rudolf P. Bohm^c, David Wang^{a,b,**}

^a Departments of Pathology & Immunology, Washington University School of Medicine, St. Louis, MO, USA

^b Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, MO, USA

^c Divisions of Veterinary Medicine and Comparative Pathology, Tulane National Primate Research Center, Covington, LA, USA

^d Nestle Purina, St. Louis, MO, USA

ARTICLE INFO

Keywords:

Gastrointestinal virome
Biogeography
Chronic diarrhea
Bacteriophages

ABSTRACT

The composition of gastrointestinal tract viromes has been associated with multiple diseases. Our understanding of virus communities in the GI tract is still very limited due to challenges in sampling from different GI sites. Here we defined the GI viromes of 15 rhesus macaques with chronic diarrhea. Luminal content samples from terminal ileum, proximal and distal colon were collected at necropsy while samples from the rectum were collected antemortem using a fecal loop. The composition of and ecological parameters associated with the terminal ileum virome were distinct from the colon and rectum samples; these differences were driven by bacteriophages rather than eukaryotic viruses. The six contigs that were most discriminative of the viromes were distantly related to bacteriophages from three different families. Our analysis provides support for using fecal loop sampling of the rectum as a proxy of the colonic virome in humans.

1. Introduction

The viral community in the primate gastrointestinal (GI) tract plays an important role in health (Virgin, 2014). The mammalian virome (Virgin et al., 2009) includes viruses that infect eukaryotic cells (eukaryotic virome), bacteriophages that infect bacteria (bacterial virome), viruses that infect archaea (archaeal virome), virus-derived genetic elements in host chromosomes as well as viruses from the diet. Because viruses are the most abundant and fastest mutating organisms on Earth, virome analysis is challenging, and we are only beginning to characterize viromes at the sequence level with the advent of next generation sequencing technology (Virgin, 2014). During early infant development the virome is highly dynamic (Lim et al., 2016). In contrast to bacterial communities, which are similar between housemates, gut viromes from healthy individuals have highly personalized bacteriophage populations, and adult viromes exhibit significant diversity between individuals (Zarate et al., 2018). However, biologically

unrelated individuals share a core set of bacteriophages that might define the normal development of the virome (Manrique et al., 2016; Zhao et al., 2017a). Alterations in the gut virome composition during infancy have been associated with pediatric disorders and the onset of autoimmune disease in later life (Reyes et al., 2015; Zhao et al., 2017a). Therefore understanding the composition and functionality of the virome as well as the interplay between virome and other components of the meiofauna (e.g. bacteria, archaea, fungi, protists and metazoans) could potentially lead to a mechanistic understanding of disease development and offer new approaches to disease prevention and therapeutic interventions.

The human GI tract is a complex system that starts at the oral cavity, continues through the stomach and intestines, and finally ends at the anus. Chemical and physical conditions as well as environmental influences and dietary preferences play an important role in determining the composition and density of the bacterial microbiota (Zarate et al., 2018). Studies of the bacterial microbiome have revealed

* Correspondence to: Department of Neuroscience, Washington University School of Medicine, 660S. Euclid Ave., Box 8108, St. Louis, MO 63110, USA.

** Correspondence to: Departments of Molecular Microbiology and Pathology & Immunology, Washington University School of Medicine, 660S. Euclid Ave., Box 8230, St. Louis, MO 63110, USA.

E-mail addresses: gzhao@wustl.edu (G. Zhao), davewang@wustl.edu (D. Wang).

¹ Current address, Department of Neuroscience, Washington University School of Medicine, St. Louis, Missouri, U.S.A.

² Current address: Vir Biotechnology, San Francisco, CA 94158, USA.

biogeographic variation along both the longitudinal and transverse axes of the digestive tract in humans and diverse animal models (Eckburg et al., 2005; Gu et al., 2013; Li et al., 2017; Looft et al., 2014; Stearns et al., 2011; Yasuda et al., 2015). In macaques, it has been shown that stool composition is correlated highly with the colonic lumen and mucosa suggesting that fecal sampling is a reasonable proxy for investigating the colonic bacterial microbiome (Yasuda et al., 2015). Much less is known about the biogeographic variation of viromes. Because virome composition and diversity does not necessarily directly correlate with that of the bacterial microbiome (Lim et al., 2015; Norman et al., 2015) and viruses can interact with the host immune system and shape the gut microbial community (Lusiak-Szelachowska et al., 2017; Manrique et al., 2017), it is important to understand the virome composition and variation along the GI tract. In the oral cavity, the viral composition of subgingival and supragingival biofilms differed from that of saliva (Ly et al., 2014). Kim and Bae (2016) compared the colonic luminal vs. mucosal virome in mice and they found that the mucosal virome significantly differed from the luminal virome in low-fat diet-fed lean mice. Currently it is unclear whether the virome changes along the longitudinal axis of the GI tract and whether the fecal virome can accurately represent viromes at internal GI sites.

Our study aimed to answer these questions by characterizing the virome at multiple distinct sampling sites (terminal ileum, proximal colon, distal colon and rectum) within the GI tract of 15 rhesus macaques with chronic diarrhea. Our analysis suggested that the virome of the terminal ileum was distinct from that of the colon and rectum and identified multiple viral contigs that were discriminative between the terminal ileum virome and viromes from other GI sites. Samples collected from the proximal colon, distal colon and rectum were generally statistically indistinguishable from each other, suggesting that fecal sampling from the rectum is a non-invasive method that accurately assesses the colonic virome.

2. Results

2.1. The enteric virome of rhesus macaques

Luminal content samples from a single small intestinal site (terminal ileum) and three large intestinal sites (proximal colon, distal colon, and rectum) were collected from 15 rhesus macaques with chronic diarrhea. Samples were collected directly from the lumen at the proximal and distal colonic sites at necropsy whereas samples from the rectum were collected antemortem using a fecal loop. We prepared libraries from combined DNA and RNA isolated from virus-like particle (VLP) preparations from these samples. Shotgun sequencing was successfully carried out on 55 samples using the paired-end 2×250 nt Illumina MiSeq platform. We obtained a mean of 1.14 ± 0.688 million sequences per sample, of which a mean of $54.3\% \pm 15.1\%$ were of high quality (Dataset S1). Deduplicated sequences were analyzed using VirusSeeker (Zhao et al., 2017b) to detect bacteriophage and eukaryotic viral sequences, which accounted for 0.09–36.2% of deduplicated sequences. There were no significant differences in the total number of sequences (Supl Fig. 1A) or the number of quality-controlled unique sequences (Supl Fig. 1B). However, terminal ileum luminal samples had a significantly lower percentage of eukaryotic viral sequences than distal colon and fecal loop samples from the rectum (Supl Fig. 1C, $P = 0.0006$ and 0.0019 respectively, Kruskal-Wallis test with Dunn's multiple comparisons test). Terminal ileum luminal samples also had a significantly lower percentage of bacteriophage sequences than proximal colon and distal colon samples (Supl Fig. 1D, $P = 0.0017$ and 0.0303 respectively, Kruskal-Wallis test with Dunn's multiple comparisons test). No differences in percentages of either eukaryotic viral sequences or bacteriophage sequences between luminal content samples collected at proximal colon, distal colon or rectum were observed.

We used two complementary approaches to compare internal viromes between different GI sites as described previously (Zhao et al.,

2017a). Briefly, bacteriophage and eukaryotic viral reads were detected using VirusSeeker (Zhao et al., 2017b) and viral abundances were normalized to total deduplicated sequences (referred to as “relative abundance” herein) from each sample and used for statistical comparison of viromes. This approach is referred to as “read-based analysis”. We also performed contig-based analysis. Sequences from each sample were individually assembled using Newbler and Phrap in a two-step-assembly approach to minimize the possibility of chimeric sequence formation (Zhao et al., 2017b). Bacteriophage and eukaryotic viral contigs were identified using VirusSeeker and were pooled and deduplicated to create a reference contig database. Individual sequencing reads from each sample were aligned to the contig database using FR-HIT (Niu et al., 2011) and a normalized matrix of the number of reads per kbp of contig sequence per million raw reads per sample (RPKM) of all the viral contigs was calculated to create a viral contig abundance matrix. This viral contig abundance matrix was used to measure α -diversity (within sample diversity) and β -diversity (between sample diversity) of the virome (Reyes et al., 2015) and to perform random forests analysis.

A total of 8990 contigs ≥ 1000 bp in length were assembled, of which 6775 were unique after deduplication. In total, 3763 unique contigs were classified as viral by VirusSeeker (Zhao et al., 2017b). The longest contig (35,012 bp) shared the greatest sequence similarity (48% identity over 552 amino acid, e value = $3e-160$) to the putative terminase large subunit of Lactobacillus virus LLKu in the family *Siphoviridae*. To identify circular full length viral genomes, we determined whether the 5' and 3' ends of a contig overlapped by at least 10 bases with $\geq 98\%$ nucleotide identity; for such contigs the overlapping sequences were trimmed, and the contig was considered a circular full-length genome. A total of 15 potential complete circular genomes were obtained. The longest of these had a length of 6208 bp after trimming and shared the highest sequence similarity (24% identity over 531 aa, e value = $2e-21$) to the major capsid protein VP1 of Parabacteroides phage YZ-2015b in the *Microviridae*. Interestingly we identified 118 contigs that shared highest sequence similarity with crAssphage, a bacteriophage initially identified to be highly abundant in human fecal metagenomes (Dutilh et al., 2014). The longest contig (21,398 bp) was most closely related to the CrAssphage sp. isolate ctcc615 with 90,023 bp complete genome. The results demonstrate that the sample preparation and sequencing depth employed sufficed to generate multiple complete viral genomes in the samples, suggesting that we have robust coverage of at least the more abundant annotatable viruses in the samples.

2.2. The enteric virome of the ileum is distinct from that of the colon and rectum

We measured virome β diversity using the Hellinger distance (a measure calculated to quantify the similarity between two viromes) metric on the RPKM matrix as described (Reyes et al., 2015). To measure the differences between viromes of the terminal ileum, proximal colon, distal colon and rectum within the same animal Hellinger distances were computed between viromes of samples collected at different GI site for a given animal (intra-animal). By Hellinger distance, the virome of the terminal ileum was equidistant to the viromes of the proximal colon, distal colon and rectum (Fig. 1A). However the virome of the proximal colon was more similar to the virome of distal colon than to the virome of terminal ileum (Fig. 1B, $p = 0.0147$). The virome of the distal colon was more similar to the viromes of proximal colon and rectum (Fig. 1C, $p = 0.0027$ and 0.0054 respectively, Kruskal-Wallis test with Dunn's multiple comparisons test) than to the virome of terminal ileum.

We assessed the extent to which the rectum virome community reflected the ileum and colonic viromes by measuring the Pearson correlation of contig relative abundances between rectum and other GI sites. Rectum virome composition was significantly correlated with that

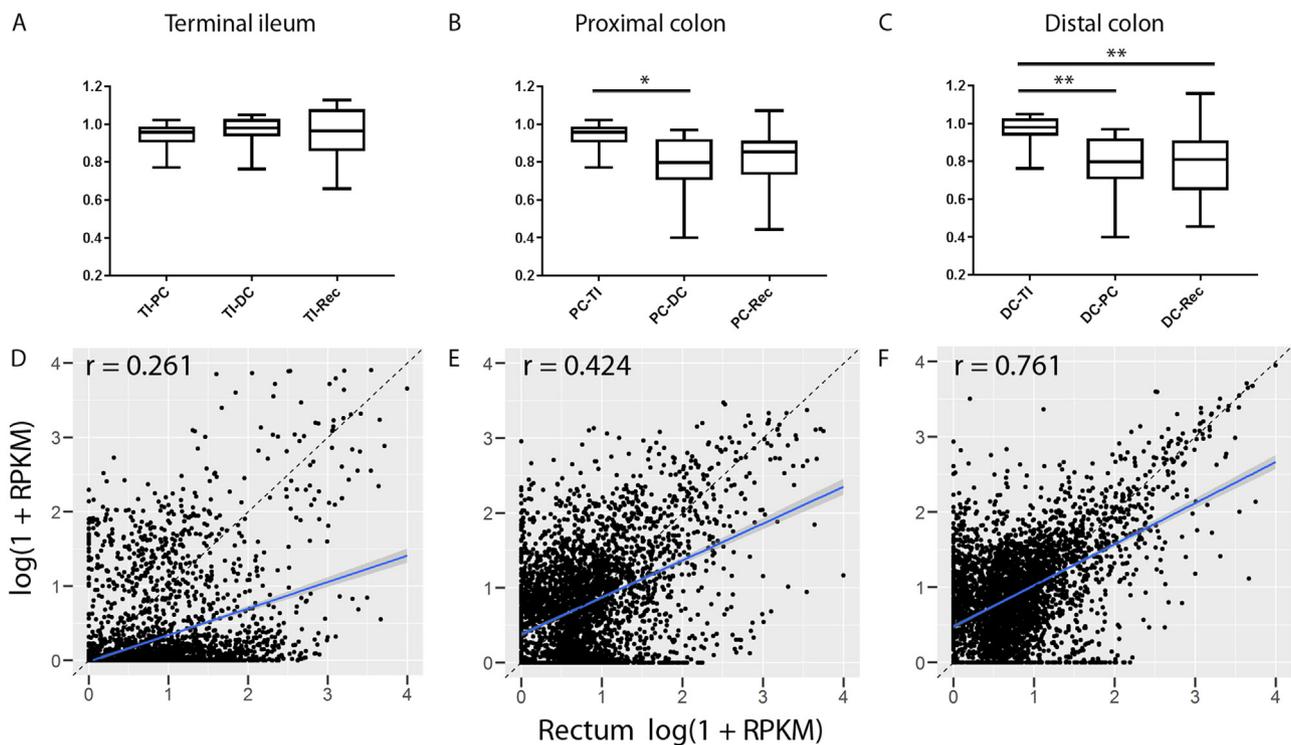


Fig. 1. Virome of terminal ileum is distinct from that of the proximal colon, distal and rectum. Mean \pm SD values for pairwise Hellinger distance-based β -diversity measurements between different GI sites within individual animals (intra-animal) are shown for (A) between terminal ileum and proximal colon, distal colon as well as rectum (B) between proximal colon and terminal ileum, distal colon as well as rectum; (C) between distal colon and terminal ileum, proximal colon as well as rectum. Differences between groups were considered statistically significant if $p < 0.05$ using nonparametric Kruskal–Wallis one-way ANOVA with Dunn's multiple comparisons test. * $p \leq 0.05$; ** $p \leq 0.01$; **** $p \leq 0.0001$. The average relative abundance of a viral contig across 15 animals for terminal ileum (D), proximal colon (E) and distal colon (F) were plotted against that of the rectum. Pearson correlation of contig relative abundances was calculated between rectum and all other sites. TI: terminal ileum; PC: proximal colon; DC: distal colon; Rec: rectum.

of the terminal ileum (Fig. 1D, $r = 0.261$, $p < 2.2 \times 10^{-16}$), proximal colon (Fig. 1E, $r = 0.424$, $p < 2.2 \times 10^{-16}$) as well as distal colon (Fig. 1F, $r = 0.761$, $p < 2.2 \times 10^{-16}$). However it was most similar to the virome composition of distal colon and most distant to that of the terminal ileum (comparing $r = 0.761$ vs. 0.261).

Comparisons were also made between viromes of individual animals (inter-animal comparison) sampled from the same GI-site (Supl Fig. 2A). The inter-animal variation was significantly lower (adjusted $P < 0.0001$, Kruskal–Wallis test with Dunn's multiple comparisons test) for the terminal ileum viromes than for the proximal colon, distal colon or rectum viromes suggesting that viromes of terminal ileum were more similar between animals than viromes at all other sites.

Taken together these results suggested that the virome of terminal ileum luminal samples was different from that of the samples collected at proximal colon, distal colon and rectum.

2.3. Bacteriophage virome of the ileum is distinct from that of the colon and rectum

Consistent with a previous report (Handley et al., 2016), bacteriophages of the *Caudovirales* order (*Myoviridae*, *Podoviridae* and *Siphoviridae* family) and *Microviridae* family were the most abundant viral taxa identified in samples collected at all four GI sites in the rhesus macaque (Fig. 2A). We performed principal coordinate analysis (PCoA) using Bray–Curtis dissimilarity distance to measure the differences between the viromes of different GI sites. Using the relative abundance matrix at the viral family level from read-based analysis suggested that the major factor contributing to the differences between viromes was the sample collection site which accounts for 31% of the variation. Luminal content samples collected at proximal colon, distal colon and rectum were clustered together, but separated from a majority of

samples collected at the terminal ileum (Fig. 2B). The same analysis using the contig abundance matrix from contig-based analysis reached the same general conclusion (Fig. 2C). In addition, species accumulation curves demonstrated that the rate of acquisition of new bacteriophage contigs in the proximal colon, distal colon and rectum samples were similar to each other, and they rapidly outpaced new taxa acquisition in terminal ileum samples (Fig. 2D).

To determine which bacteriophage taxa contributed to the differences between the viromes we compared viral sequence relative abundance, Shannon diversity and richness of viromes using both read-based analysis and contig-based analysis. Generalized linear mixed-effects models that accounts for correlations between different GI sites were used to test whether different GI sites were significantly associated with the change of characteristics of the virome. Tukey's HSD (honest significant difference) test was used post-hoc to identify which pairs of viromes were different. The relative abundance of sequences assigned to *Microviridae*, *Myoviridae* and *Siphoviridae* was significantly higher in the virome of proximal colon than that of the terminal ileum (Table 1, Fig. 3A, B, D). Similarly, compared to the viromes of terminal ileum, the relative abundance of sequences assigned to *Microviridae* and *Siphoviridae* was significantly higher in the virome of the distal colon (Table 1, Fig. 3A, D), and the relative abundance of sequences assigned to *Microviridae* was significantly higher in the viromes of the rectum (Table 1, Fig. 3A). No differences in relative abundance for any of the virus families were observed between samples collected from the proximal colon, distal colon and rectum.

Viromes of the proximal colon, distal colon and rectum all exhibited higher richness than that of the terminal ileum for all the viral families tested (Table 1, Fig. 3E–H). No differences in richness were observed between samples collected at proximal colon, distal colon and rectum. Similar results were obtained using contig-based analysis for all viral

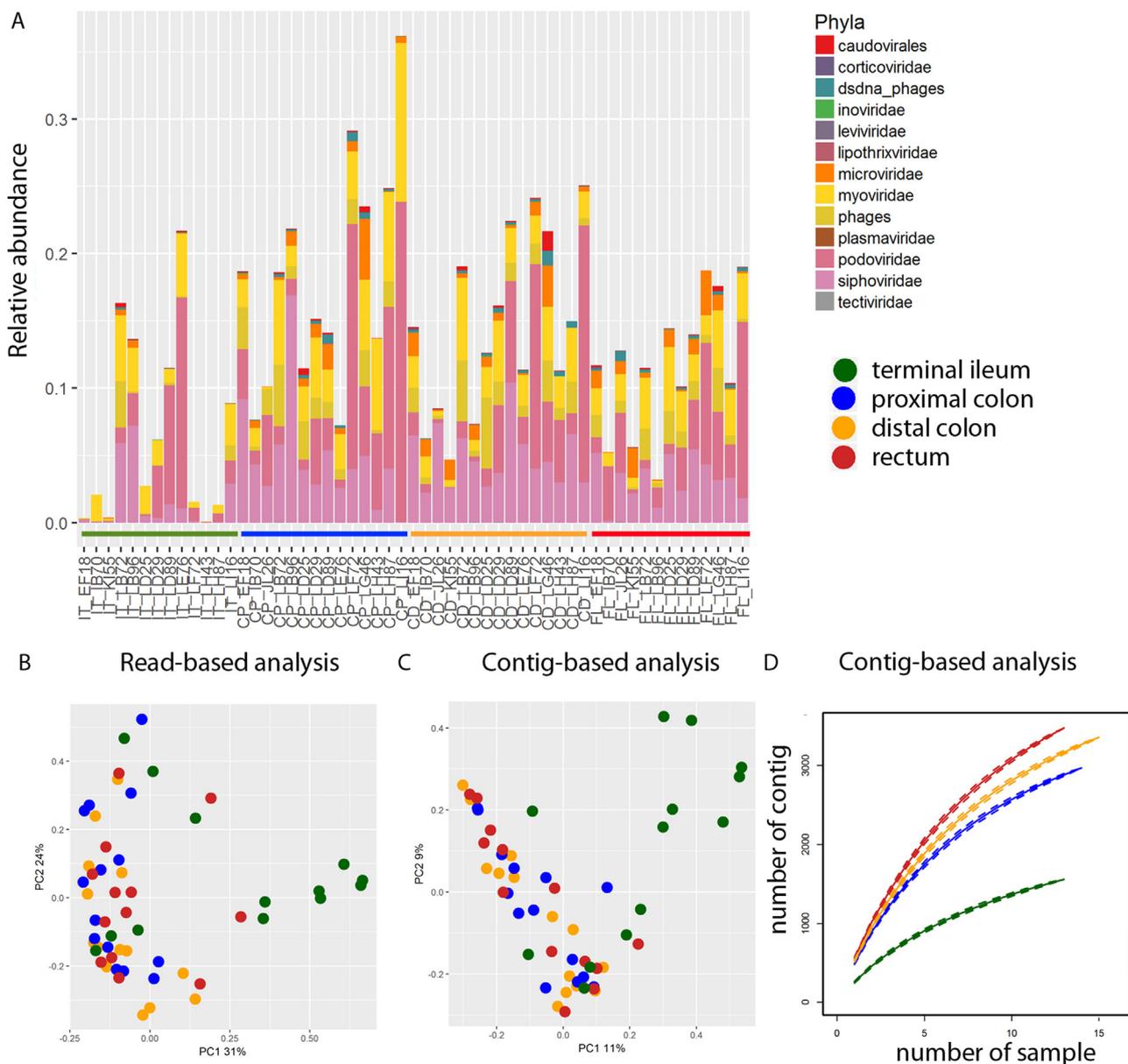


Fig. 2. Biogeographic influences on macaque gut virome composition. (A) Family-level relative abundance of intestinal bacteriophage virome of 15 rhesus macaques. Principal coordinate analysis (PCoA) of all samples by Bray–Curtis dissimilarity distance using B) read-based analysis and (C) contig-based analysis. (D) Species accumulation curves of bacteriophage richness versus an increasing number of samples based on assembled contigs.

families (Suppl. Fig. 3A-D).

In summary, the bacteriophage virome of the ileum was distinct from that of the colon and rectum whereas no differences were observed between the virome of rectum and that of the colon by any of the

metrics.

Table 1

Comparison of viral sequence relative abundance, read-based analysis of Shannon diversity and richness between viromes of terminal ileum (TI), proximal colon (PC), distal colon (DC) and rectum (Rec).

Virus	Measurement	TI-PC	TI-DC	TI-Rec	PC-DC	PC-Rec	DC-Rec
Microviridae	Relative abundance	0.0317	0.0357	0.0444	NS (1.00)	NS (0.999)	NS (0.993)
	Richness	0.0293	0.00580	< 0.001	NS (0.622)	NS (0.150)	NS (0.720)
Myoviridae	Relative abundance	0.00121	NS (0.640)	NS (0.741)	NS (0.0615)	NS (0.0713)	NS (1.00)
	Richness	0.0245	< 0.001	< 0.001	NS (0.306)	NS (0.0608)	NS (0.793)
Podoviridae	Relative abundance	NS (0.362)	NS (0.910)	NS (0.923)	NS (0.674)	NS (0.571)	NS (0.998)
	Richness	0.00319	< 0.001	< 0.001	NS (0.226)	NS (0.269)	NS (1.00)
Siphoviridae	Relative abundance	< 0.001	< 0.001	NS (0.259)	NS (1.00)	NS (0.298)	NS (0.0895)
	Richness	0.0279	0.00171	0.00154	NS (0.805)	NS (0.800)	NS (1.00)

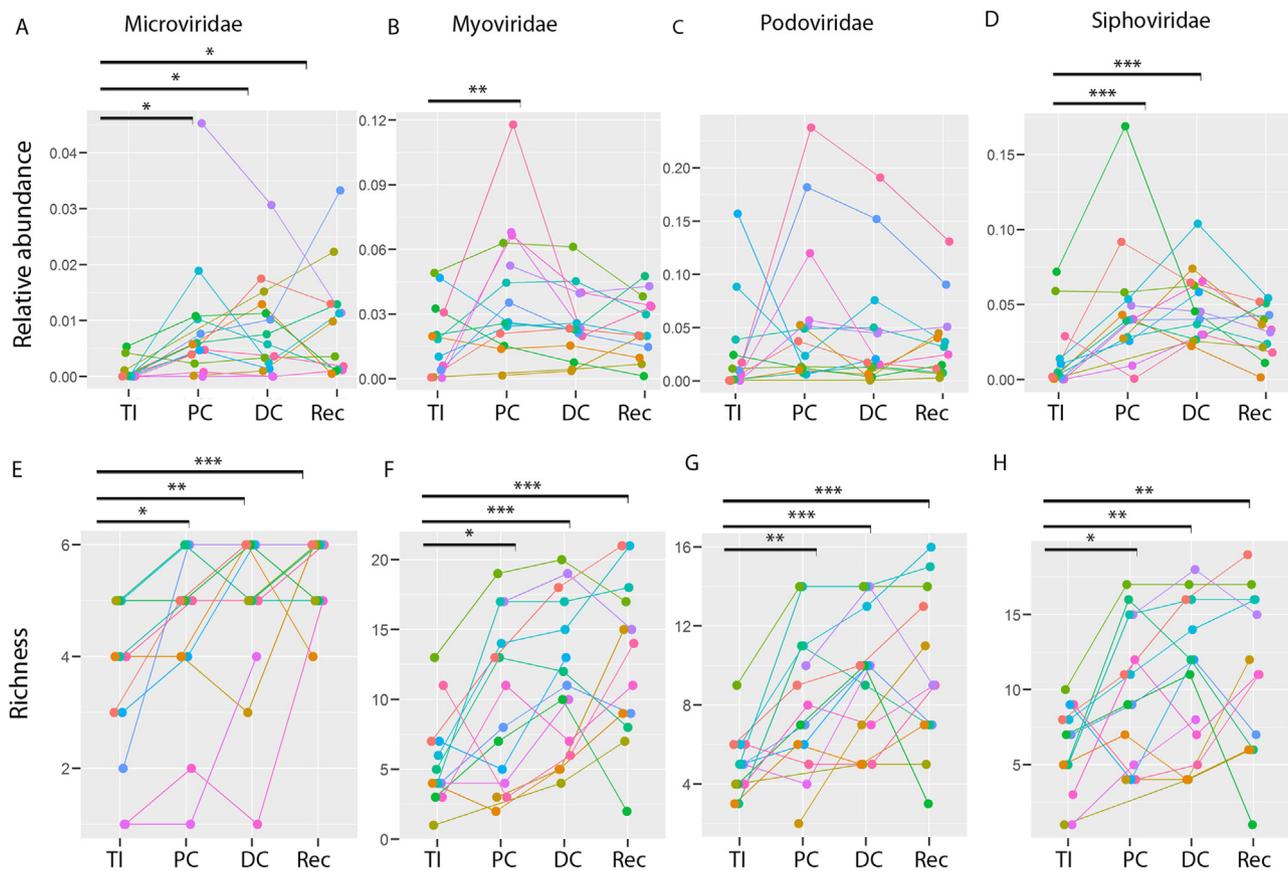


Fig. 3. Alterations in viral community composition along the GI tract. The relative abundance (top) and richness (bottom) of (A, E) *Microviridae*, (B, F) *Myoviridae*, (C, G) *Podoviridae* and (D, H) *Siphoviridae* were compared between terminal ileum, proximal colon, distal colon and rectum of the same animals using read based analysis. Differences between groups were considered statistically significant if $p \leq 0.05$ using generalized linear mixed-effects models. Tukey's HSD (honest significant difference) test were used as post-hoc analysis to identify which pairs of viromes were different. * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$. TI: terminal ileum; PC: proximal colon; DC: distal colon; Rec: rectum. Each color represents a different animal.

2.4. Identification of contigs that discriminate viromes of the terminal ileum from those of proximal colon, distal colon and rectum

We used random forests algorithm to identify contigs that can distinguish the viromes of different GI sites. The initial model using all 3763 contigs to classify 4 different GI sites performed poorly as it had an out-of-bag (OOB) estimate of error rate of 81.82%. It did classify terminal ileum samples with relative high accuracy (OOB error rate of 23.1%) but failed to distinguish samples collected at proximal colon, distal colon and rectum (Suppl. Fig. 4A). Samples from the terminal ileum were clustered together separate from other sample types in a multidimensional scaling (MDS) plot of the proximity matrix from the random forests model (Fig. 4A). To try and improve the performance of the random forests classifier, we used a method that fits random forests models by iteratively discarding the contigs with the smallest feature importance score (Diaz-Uriarte and Alvarez de Andres, 2006) to identify GI site-discriminative contigs. This analysis was repeated 800 times. The average out-of-bag posterior probability of each sample belonging to each class was calculated. As before, samples of terminal ileum could be accurately classified (Suppl. Fig. 4B) but samples from other GI sites could not be distinguished (Suppl. Fig. 4C, D, E). These results suggest that the viromes of proximal colon, distal colon and rectum were not distinguishable from each other, but they could be distinguished from that of the terminal ileum.

We measured the frequency of a contig being selected as the discriminative contig in the 800 bootstrap iterations to quantify sampling variability and confidence in discriminative-feature identification as described (Pepe et al., 2003). Fig. 4B shows the selection probability of

each contig for the first 200 most-frequently selected contigs. We defined those contigs with a selection probability of three standard deviations away from the mean selection probability ($P < 10^{-3}$) as the most-discriminative contigs. By this criterion, we identified 6 contigs that were most discriminative for different GI sites (Fig. 4B), and these contigs were ranked by the selection probability in the bootstrap samples (Table 2). To compare the discriminative power of the 6 contigs directly with all 3763 contigs we treated samples from proximal colon, distal colon and rectum as one group and tested the ability of the contigs to distinguish samples collected from these sites (large intestine) and samples collected from terminal ileum (small intestine). A random forests model using all 3763 contigs had an OOB estimate of error rate of 7.27% and the area under the receiver operating characteristic curve (AUC) of 0.890 whereas a random forests model using the 6 GI site-discriminative contigs had similar performance with an OOB estimate of error rate of 5.45% and AUC of 0.890 (Fig. 4C). MDS plot of the proximity matrix (Fig. 4D) from the random forest model using the 6 GI site-discriminative contigs achieved similar separation of samples as using all 3763 contigs. Fig. 4E showed a heat map of the abundance of the 6 GI site-discriminative contigs.

Next, we used generalized linear mixed-effects models to test whether the abundance of the GI site-discriminative contigs were different between different GI sites. The abundance of contig 1, 4 and 5 (Fig. 5A, D, E and Table 2) was significantly higher in the terminal ileum than all other GI sites. The abundance of contig 2 and 6 (Fig. 5B, F and Table 2) was significantly lower in the terminal ileum than in the proximal colon. The abundance of contig 2 was also significantly lower in the terminal ileum than that of the distal colon (Fig. 5B and Table 2). No

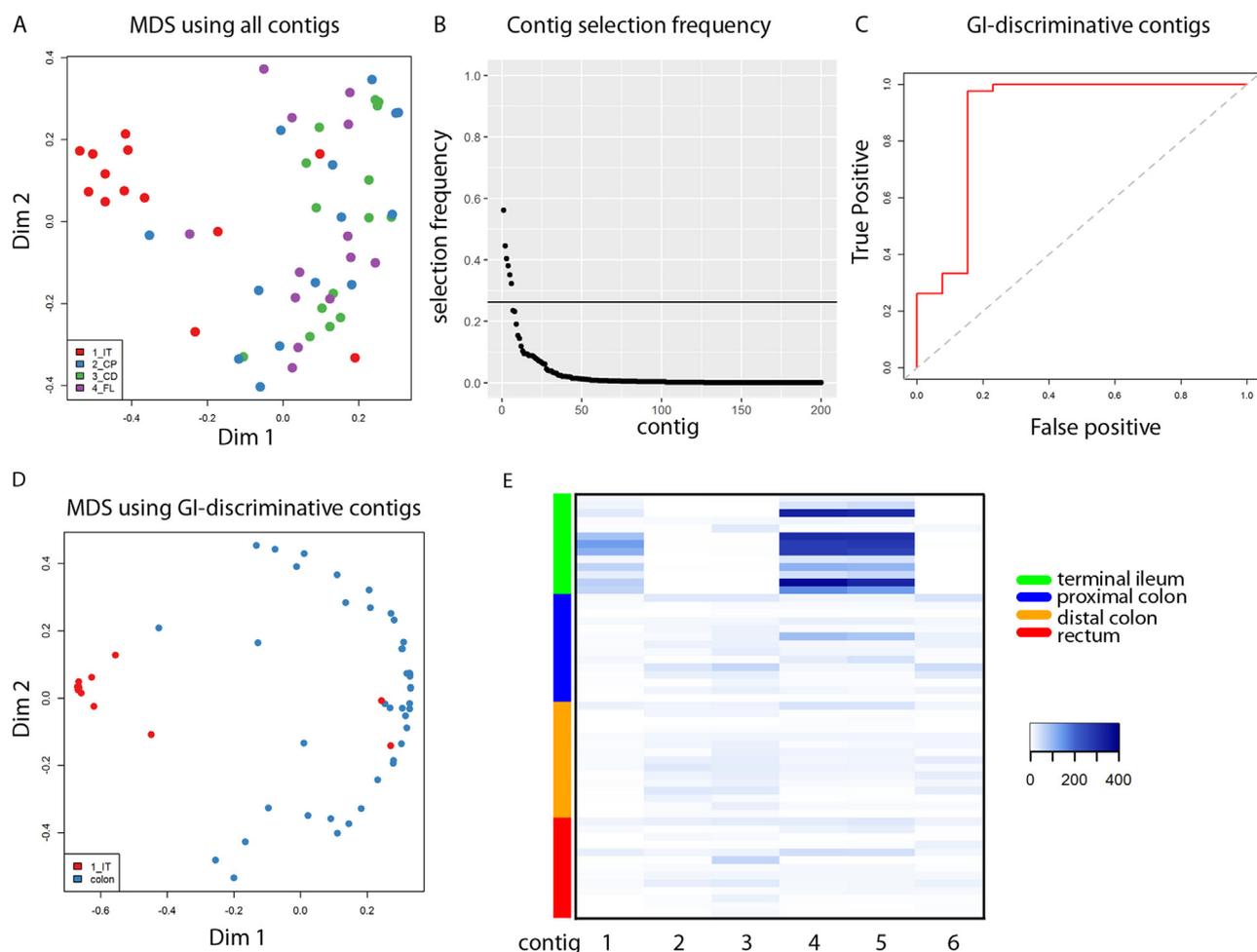


Fig. 4. Identification of GI-site discriminative contigs. (A) Multidimensional scaling (MDS) plot of the proximity matrix from the random forests model using all 3763 contigs to classify 4 different GI sites. (B) The selection probability of the first 200 most-frequently selected contigs. (C) The area under the receiver operating characteristic curve (AUC) of a random forests model using the 6 GI site-discriminative contigs. (D) MDS plot of the proximity matrix from the random forest model using the 6 GI site-discriminative contigs to classify terminal ileum vs other GI sites. (E) Heat map of the abundance of 6 GI site-discriminative contigs. Each row represents a sample. Each column represents a contig, and contigs are arranged according to decreasing selection probability.

significant difference in contig abundance were observed between proximal colon, distal colon or rectum for any of the GI site-discriminative contigs.

All 6 discriminatory contigs shared sequence similarity with reference sequences derived from both bacteria and bacteriophages. Contig 4 and 5 were most closely related. They were obtained from two different animals and partially overlapped sharing 97% nt sequence identity over 1248 nt. Contig 4 and 5 shared 83–86% nt sequence identity with contig 1. All three contigs were most closely related to bacteria in the *Proteobacteria* phylum (88–95% nt sequence identity, e value = 0) and *Dickeya* phage phiDP10.3 (82–87% nt sequence identity) (Table 2). Contig 2 and 3 were obtained from assembling sequences from two different animals but overlapped 2352 nt with 99.9% nucleotide identity at the end of the contigs suggesting that they were derived from the same virus. Both contigs shared highest sequence similarity with *Bacteroides fragilis* strain HMW 615 transposon CTnHyb (95% nt identity, e value = 0) and *Staphylococcus* phage UPMK_1 (71% nt identity, e value = $1E-24$). Contig 6 was most closely related to Uncultured bacterium clone PITZ_12F_Contig_20 (91% nt identity, e value = $9E-103$) and was distantly related to Planktothrix phage PaV-LD (36% aa identity, e value = $2E-08$).

Since Contig 2 and 3 seemed to come from the same virus we performed further assembly using unique sequences from samples of all GI sites from the two animals from which Contig 2 and 3 were obtained. A

contig of 12294 bp was obtained which included complete sequences of both Contig 2 and 3. This contig shared the highest sequence similarity to the complete genome of *Bacteroides salanitronis* DSM 18170 (CP002530.1) covering 63% of the contig with 63–77% nt identity. The region of the bacteria genome with sequence similarity to the contig included a putative mobilization protein, a phage integrase family protein and a conjugate transposon protein suggesting that this region is likely an unannotated mobile element although the precise nature of the mobile element needs further characterization. Combined with the significant similarity of this large contig to known, annotated phage sequences, these data suggest that the sequences detected are derived from bacteriophage (albeit we cannot definitively determine whether it is in the form of free virus or as a prophage). Contig 6 shared 91% (1800/1980) nt identity over the full length to this 12294 bp contig suggesting that Contig 6 was derived from a closely related virus.

2.5. No significant difference in eukaryotic viromes of the terminal ileum, proximal colon, distal colon and rectum

Eukaryotic viruses from nine different families were detected (Fig. 6A). *Alloherpesviridae*, *Astroviridae*, and *Reoviridae* were excluded from further ecological analysis because viruses in these families were detected in ≤ 5 samples. We observed that the relative abundance of sequences assigned to the *Circoviridae*, *Picobirnaviridae*, *Picornaviridae*

Table 2
GI site-discriminative contigs.

	Contig1	Contig2	Contig3	Contig4	Contig5	Contig6
Length (bp)	1409	4880	7050	1647	1496	1978
Selection Frequency	0.56125	0.445	0.40375	0.38125	0.35125	0.3225
TI-PC*	< 0.001	0.0309	NS (0.242)	< 1e-04	< 1e-04	0.0126
TI-DC*	< 0.001	0.0169	NS (0.622)	< 1e-04	< 1e-04	NS (0.576)
TI-Rec*	0.00184	NS (0.699)	NS (0.352)	0.000117	< 1e-04	NS (0.942)
PC-DC*	NS (1.00)	NS (0.999)	NS (0.891)	NS (1.00)	NS (1.00)	NS (0.287)
PC-Rec*	NS (1.00)	NS (0.490)	NS (0.996)	NS (1.00)	NS (1.00)	NS (0.0895)
PC-Rec*	NS (1.00)	NS (0.303)	NS (0.968)	NS (1.00)	NS (1.00)	NS (0.911)
Most closely related Bacteriophage	Dickeya phage phiDP10.3 clone pD10.contig.26_1	Staphylococcus phage UPMK_1	Staphylococcus phage UPMK_1	Dickeya phage phiDP10.3 clone pD10.contig.26_1	Dickeya phage phiDP10.3 clone pD10.contig.26_1	Planktothrix phage PaV-LD]
Accession ID	KM209255.1	MG543995.1	MG543995.1	KM209255.1	KM209255.1	YP_004957396.1
Length	5269 (nt)	152,788 (nt)	152,788 (nt)	5269 (nt)	5269 (nt)	197 (aa)
% identity (36%) aa	1158/1418 (82%)	199/280 (71%)	199/280 (71%)	1224/1412 (87%)	1293/1510 (86%)	57/159
e value	0	1.00E-24	1.00E-24	0	0	2.00E-08
Most closely related Bacteria	Rhizobacter gummiphilus strain NBRC 109400 chromosome	Bacteroides fragilis strain HMW 615 transposon CTnHyb	Bacteroides fragilis strain HMW 615 transposon CTnHyb	Actinobacillus equuli subsp. equuli strain 19392	Actinobacillus equuli subsp. equuli strain 19392	Uncultured bacterium clone PITZ_12F.Contig_20
Accession ID	CP024645.1	KJ816753.1	KJ816753.1	CP007715.1	CP007715.1	KU548052.1
Length	6398100 (bp)	131471 (bp)	131471 (bp)	2431533 (bp)	2431533 (bp)	274 (bp)
% ID (91%)	1231/1396 (88%)	1382/1719 (80%)	1383/2027 (68%)	1332/1399 (95%)	1426/1495 (95%)	250/274
e value	0	0	0	0	0	9.00E-103

Note: TI-CP, TI-CD, TI-FL, CP-CD, CP-FL, CD-FL: Comparison of relative abundance of corresponding contigs between each pairs of viromes using Post-hoc Tukey's HSD (honest significant difference) test. TI: terminal ileum; PC: proximal colon; DC: distal colon; Rec: rectum.

and Smacovirus group was lower in the terminal ileum samples than in the samples collected at proximal colon, distal colon and rectum whereas no difference were observed for the *Adenoviridae* and *Parvoviridae* (Fig. 6). Comparison using generalized linear mixed-effects models which accounts for correlation of viral abundance at different GI sites and Tukey's HSD test as post-hoc analysis did not identify any statistical significant differences in pair-wise comparisons between sites for any of the viruses. The prevalence, based solely on presence or absence, of sequences assigned to each viral family was not significantly different between terminal ileum and rectum samples by McNemar's test which is a statistical test used on paired nominal data. Based on these analyses, we conclude that no differences in specific eukaryotic viruses or viral taxa were observed between the different anatomic sites. In addition, no differences in eukaryotic viral richness or diversity were observed between the sites.

3. Discussion

In this study, we investigated the virome of multiple sites along the GI tract of rhesus macaques. Our results demonstrate that the rectum virome community is a good proxy for the large intestinal virome, but it is distinct from the virome of the terminal ileum. This study thus provides the first quantitative measure of the relationships between the viral communities along the longitudinal axis of the nonhuman primate GI tract.

The mammalian small intestine and large intestine have distinct structure features and immune organs. For example the small intestine is the primary site for reabsorption of bile and has circular folds, villi and Peyer's patches which are absent in the large intestine. These differences, as well as differences in enzyme secretion and pH, lead to physiological variations along the length of the GI tract which in turn can influence bacterial community composition. The ileum is the terminal portion of the small intestine which is distinct from colon. Several bacterial microbiome studies have shown that ileal contents showed reduced bacterial microbiome diversity (richness and abundance) when compared with fecal samples or samples collected in the large intestine (Gu et al., 2013; Li et al., 2017; Looft et al., 2014).

Bacteria in the colon account for approximately 70% of all bacteria in the human body (Hillman et al., 2017). The lower bacterial numbers in the small intestine have been attributed to low pH and short transit time (Berg, 1996). The colon is the main site for the bacterial fermentation of non-digestible food components such as soluble fiber and reabsorption of water. The change of pH, low cell turnover rate, low redox potential, and the long transit time could all contribute to the higher biodiversity of bacteria in the colon. In this study we determined that the relative abundance, diversity and richness of bacteriophages were low in the ileum compared to the large intestine. Similarly, the bacterial microbiome of the ileum in healthy people and mammals has been reported to have lower diversity and richness compared to the large intestine. Samples used in this study were obtained from rhesus macaques with chronic diarrhea. It has been shown that both acute and chronic diarrhea are related to a dysregulation of intestinal homeostasis and gut bacterial microbiome composition (Pop et al., 2014; Scaldaferrri et al., 2012). Chronic diarrhea has also been associated with differences of virome composition between rhesus macaque with diarrhea and healthy controls (Kapusinszky et al., 2017) with the caveat that fecal samples from sick animals whereas rectal swabs from healthy animals were used in the study. In this cohort of macaques with chronic diarrhea, low diversity in the ileum was observed compared to the colon. It is possible that similar trends also characterize the enteric tract of healthy macaques and primates in general. However, as differences in the virome between healthy animals and those with diarrhea could exist (Kapusinszky et al., 2017), it is also possible that our observations are specific to animals with diarrhea. Future studies of healthy animals are necessary to directly address this question although we note that the difficulty of obtaining such samples from multiple sites in the enteric tract from healthy controls poses significant challenges.

The relationships between viromes and different diseases are complex. In type I diabetes, no obvious bacterial microbiome changes were observed before the development of autoimmunity whereas lower diversity and richness were observed for the virome (Kostic et al., 2015; Zhao et al., 2017a). In patients with inflammatory bowel disease, bacterial microbiome diversity is lower whereas the bacteriophage virome expands (Norman et al., 2015). By contrast in HIV infected

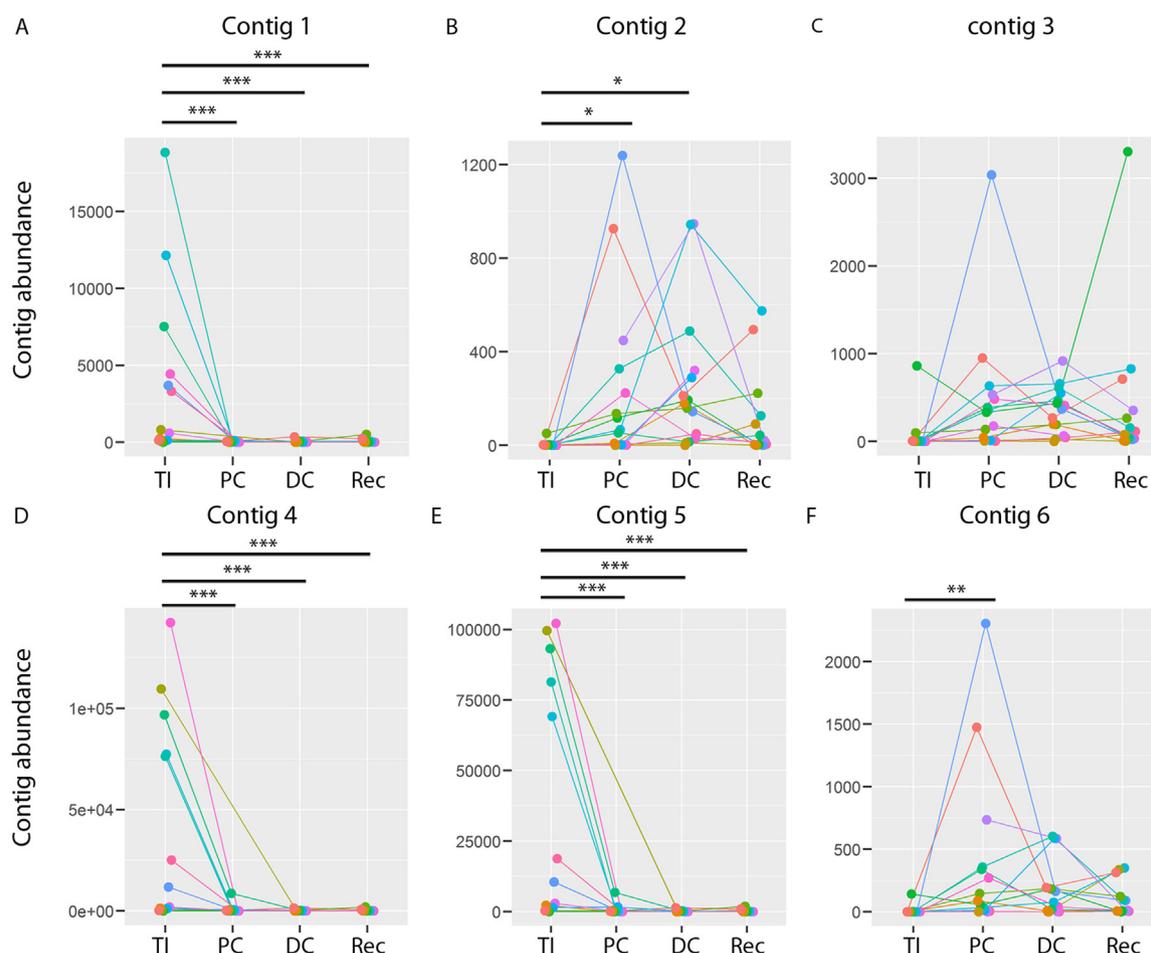


Fig. 5. Comparison of relative abundance of GI-discriminative contigs between GI sites. The relative abundance of GI discriminative contig were compared between terminal ileum, proximal colon and rectum of the same animals. Viral contig abundance were measured by RPKM. Differences between groups were considered statistically significant if $p \leq 0.05$ using generalized linear mixed-effects models. Tukey's HSD (honest significant difference) test were used as post-hoc analysis to identify which pairs of viromes were different. * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$. TI: terminal ileum; PC: proximal colon; DC: distal colon; Rec: rectum. Each color represents a different animal.

patients and SIV infected macaques, the eukaryotic virome expands, but there is no obvious change of the bacterial microbiome until the subjects are severely immune compromised (Gootenberg et al., 2017; Handley et al., 2016, 2012; Monaco et al., 2016). Currently it is unclear whether virome changes are causal or merely correlated with disease conditions. Better understanding of the virome composition under normal physiological conditions along the GI tract will help us to better understand the causal relationships between virome changes and many diseases.

A study demonstrated that interpersonal variation contributes significantly to variance between virome samples from humans (Minot et al., 2011); another study of healthy adult female monozygotic twins and their mothers has shown that viromes have even greater interpersonal variability than the bacterial microbiome (Reyes et al., 2010). In mice the bacterial microbiome of the small intestine exhibited higher inter-subject variation than that of the large intestinal and fecal samples (Gu et al., 2013). Taken together, these data would suggest that the virome of the small intestine might have even higher inter-subject variation than the large intestine and fecal samples. However in our study we observed lower inter-subject variation of viromes in the ileum comparing to that of the large intestine. Whether the differences are due to the differences in regional GI tract physiology or due to distinct relationships between viruses and bacteria in different species awaits further analysis.

In ileostomy samples from humans, the small intestine was found to

exhibit lower bacterial diversity than the colon and was highly enriched in certain members of the phylum *Proteobacteria* (Zoetendal et al., 2012). Similarly *Proteobacteria* were found to be enriched in the small intestine of mice (Gu et al., 2013). Interestingly the ileum-discriminative contigs 1, 4 and 5, which have significantly higher relative abundance in the terminal ileum than in the large intestine (Fig. 5A, D, E and Table 2), were closely related to the bacteria in the *Proteobacteria* phylum suggesting that they could represent bacteriophages that infect bacteria in the *Proteobacteria* phylum.

It is challenging to annotate bacteriophage sequences in the metagenomics data. The relationships between bacteriophage and bacteria in term of genomic sequences are complicated. Lysogenic phage can exist as integrated pro-phage in bacterial genomes. Part of bacteriophage genes can be left in the host bacteria genome when a phage excise from the host genome. On the other hand bacteriophage can carry part of host bacteria genes and serve as vectors to transfer bacteria sequences between hosts. As in many other metagenomics studies we focus on experimental enrichment of viral particles prior to sequencing. Our virus-like particle preparation employed filtration combined with lysozyme and nuclease treatment to enrich for virus sequences. All samples at different GI sites were sampled to the same total read depth. Contig assembly and annotation yielded multiple complete circular bacteriophage genomes demonstrated that the sample preparation and sequencing depth employed sufficed to generate robust coverage of at least the more abundant viruses in the samples. It has

been shown that read length significantly influences the annotation process with increased sensitivity at longer lengths (Carr and Borenstein, 2014). Short reads that shared no detectable sequence similarity with anything in the database could be assembled into longer contigs that can be annotated more accurately. Therefore contig-based analysis could reflect the viral community more accurately than read-based analysis. We performed both individual read-based analysis and contig-based analysis and both approaches reached the same conclusion demonstrating the robustness of the conclusion.

Due to difficulties of sampling, most human virome and bacterial microbiome studies use fecal samples as a proxy of the virome in the gut (Consortium, 2012; Lim et al., 2015; Monaco et al., 2016; Qin et al., 2010, 2012; Reyes et al., 2015; Yatsunenkov et al., 2012; Zhao et al., 2017a). Previous studies that have examined microbial communities directly in the colon suggest that stool microbiome composition correlated highly with the colonic and moderately with the distal small intestine microbiome (Hold et al., 2002; Yasuda et al., 2015). In this study we have shown that, similar to the bacterial microbiome, the virome of the rectum significantly correlated with the virome of the colon and moderately with the distal small intestine. These results support fecal loop sampling as an accurate representation of colonic samples for translational studies.

4. Material and methods

4.1. Animals

Fifteen rhesus macaques (*Macaca mulatta*) that presented to the breeding colony hospital a minimum of three months prior to sample collection and at the time of sample collection and met the following criteria were included in the study: intractable fluid stool (75% of daily observations since presentation with a stool consistency characterized as “soft or fluid”); Body Condition Score (BCS) of 2/5 or greater; euhydration; no other significant abnormal physical examination findings; no current antimicrobial treatment. These animals demonstrated gastrointestinal disease with no, or minimal signs of systemic illness. Samples were collected from the rectum antemortem using fecal-loop sampling. Terminal ileal contents as well as contents of proximal and distal colon were collected at necropsy.

4.2. Virome sequencing

For each sample, all available samples up to 200 mg was re-suspended in 6X volume of SM buffer and centrifuged (Finkbeiner et al., 2009). The fecal supernatant was passed through a 0.45- μ m-pore-size membrane. The filtrates were treated with Lysozyme and DNase at 37 °C for 1 h. Total RNA and DNA was extracted on a COBAS AmpliPrep instrument (Roche) according to the manufacturer's recommendation. In order to evaluate samples for both RNA and DNA viruses the total nucleic acids were randomly amplified as described previously (Finkbeiner et al., 2009; Wang et al., 2003) using barcoded primers consisting of a base-balanced 16-nucleotide-specific sequence upstream of a random 15-mer and used for NEBNext library construction (New England BioLabs). Libraries of 55 samples were successfully constructed and were multiplexed (12 samples per flow cell) on an Illumina MiSeq instrument (Washington University Center for Genome Sciences) using the paired-end 2 \times 250 protocol. Raw data have been deposited in the Sequence Read Archive (accession no. PRJNA491509).

4.3. Identification and analysis of viral-like sequences

VirusSeeker-Virome was used to detect sequences sharing nucleotide and protein level sequence similarity to known viruses. VirusSeeker was designed to remove potentially ambiguous or false-positive viral sequences and identify only sequences unambiguously classified as viral sequences. Briefly, sequences are adapter-trimmed, joined if read1 and

read 2 of paired end reads overlaps, quality controlled and dereplicated. Potential unique viral sequences were queried against the NCBI nt/nr databases, and only sequences matching exclusively to viral sequences were kept for further analysis. All sequences aligning to viruses were further classified into viral families based on the NCBI taxonomic identity of the best hit.

Absolute read counts were normalized by dividing individual taxon sequence counts by the total number of deduplicated sequence count in a sample (referred to as “relative abundance”). Richness and diversity were calculated using the *diversityresult* function of the *vegan* package. Heatmaps were generated using the *gplots* R package. Virus presence/absence graphs were generated using *ggplot2* R package.

4.4. Assembly of viral contigs and cross-contig comparison

VirusSeeker-Discovery (Zhao et al., 2017b) was used to assemble sequences into contigs and to detect contigs sharing nucleotide and protein level sequence similarity to known viruses. The resulting contigs from each sample were pooled and deduplicated using CD-HIT to cluster sequences at 95% sequence identity with 95% overlap. A total of 8990 contigs \geq 1000 bp in length and 3763 unique contigs were obtained after deduplication.

To analyze viral abundance and diversity within and between individuals, all high quality, deduplicated sequences from each sample were mapped against the assembled contigs using FR-HIT (Niu et al., 2011) at 90% identity cutoff. The crAssphage (Dutilh et al., 2014) genome were added to the 3763 unique contigs which lead to 3764 final contigs used for the analysis. A matrix of reads per Kbp of contig sequence per million reads of sample (RPKM) was generated and used for analysis of α - and β -diversity.

4.5. Viral α - and β -diversity

For α -diversity analysis, “Observed species” and the “Shannon diversity index” were calculated for each sample using *BiodiversityR* R package. To calculate α -diversity metrics associated with *Microviridae*, *Myoviridae*, *Podoviridae* and *Siphoviridae* only the rows corresponding to contigs assigned to the given taxonomy were selected from the matrix. β -Diversity analyses were performed using the RPKM abundance matrix; the Hellinger distance matrices were calculated and GraphPad Prism version 6 (GraphPad Software) was used to evaluate between animal and between sites virome sample distances.

4.6. Random forests analysis

For GI site-discriminative contig analysis, Random forests was used as a classifier for discrete variable “GISite”. A total of 800 bootstrap with 200 random trees for each bootstrap data set were generated. The out-of-bag (OOB) error rate, feature importance, and prediction from the 800 bootstrap dataset were averaged for the results presented. The *VarSelRF* R package (Diaz-Urriarte, 2007) was used for feature selection which select the minimum number of contigs capable of obtaining the same or better classification accuracy, thus minimizing the error. In the 800 bootstrap iterations and each with 200 trees the frequency of a contig being selected as discriminatory was calculated (referred to as “selection probability” herein) and the contigs were ranked by the selection probability. The higher the selection probability a contig had the higher confidence we had on the discriminative property a contig had. We defined those contigs whose selection probability was three standard deviation away from the mean selection probability ($P < 10^{-3}$) as the most-discriminative contigs for GI sites.

4.7. Statistical analysis

The *BiodiversityR* function of the *vegan* package was used to calculate diversity and richness measurements of viromes. Function

specaccum in the vegan package was used to calculate the number of species for an increasing number of animals. The *ggplot* function in the *ggplot2* package was used for plotting. The *glmmPQL* function of generalized linear mixed-effects models in the *MASS* R package was used to compare different measures between terminal ileum, proximal colon, distal colon and rectum except for number of total reads and number of unique reads which were not correlated with GI sites. Penalized quasi-likelihood (PQL) is a flexible technique that can deal with non-normal data, unbalanced design, and crossed random effects. The *glt* function of Tukey's HSD test in the *multcomp* R package was used for pairwise comparisons between GI sites. Associations between groups using number of subjects were determined using two-by-two contingency tables and significance was determined with Fisher's exact test. Pearson's product moment correlation were calculated using the *cor.test* functions in *stats* R package. Correlation plot was drawn using the *corrplot* R package. Univariate and multivariate statistical tests were performed in GraphPad Prism version 6 (GraphPad Software). P values ≤ 0.05 were considered significant.

Acknowledgments

We would like to thank Brian Koebe and Eric Martin from the High Throughput Computing Facility at the Center for Genome Sciences and Systems Biology for providing high-throughput computational resources and supports; Jessica Hoisington-Lopez from the DNA Sequencing Innovation Lab at the Center for Genome Sciences and Systems Biology for her sequencing expertise; This work was supported by National Institutes of Health Office of The Director, grants R24 OD019793, National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID) and Washington University DDRCC Grant (NIDDK P30 DK052574).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.virol.2018.10.001](https://doi.org/10.1016/j.virol.2018.10.001).

References

- Berg, R.D., 1996. The indigenous gastrointestinal microflora. *Trends Microbiol.* 4, 430–435.
- Carr, R., Borenstein, E., 2014. Comparative analysis of functional metagenomic annotation and the mappability of short reads. *PLoS One* 9, e105776.
- Consortium, H.M.P., 2012. Structure, function and diversity of the healthy human microbiome. *Nature* 486, 207–214.
- Diaz-Urriarte, R., 2007. GeneSrf and varSelRF: a web-based tool and R package for gene selection and classification using random forest. *BMC Bioinform.* 8, 328.
- Diaz-Urriarte, R., Alvarez de Andres, S., 2006. Gene selection and classification of microarray data using random forest. *BMC Bioinform.* 7, 3.
- Dutilh, B.E., Cassman, N., McNair, K., Sanchez, S.E., Silva, G.G.Z., Boling, L., Barr, J.J., Speth, D.R., Seguritan, V., Aziz, R.K., et al., 2014. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat. Commun.* 5, 4498.
- Eckburg, P.B., Bik, E.M., Bernstein, C.N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S.R., Nelson, K.E., Relman, D.A., 2005. Diversity of the human intestinal microbial flora. *Science* 308, 1635–1638.
- Finkbeiner, S.R., Holtz, L.R., Jiang, Y., Rajendran, P., Franz, C.J., Zhao, G., Kang, G., Wang, D., 2009. Human stool contains a previously unrecognized diversity of novel astroviruses. *Virology* 397, 161–167.
- Gootenberg, D.B., Paer, J.M., Luevano, J.M., Kwon, D.S., 2017. HIV-associated changes in the enteric microbial community: potential role in loss of homeostasis and development of systemic inflammation. *Curr. Opin. Infect. Dis.* 30, 31–43.
- Gu, S., Chen, D., Zhang, J.N., Lv, X., Wang, K., Duan, L.P., Nie, Y., Wu, X.L., 2013. Bacterial community mapping of the mouse gastrointestinal tract. *PLoS One* 8, e74957.
- Handley, S.A., Desai, C., Zhao, G., Droit, L., Monaco, C.L., Schroeder, A.C., Nkolola, J.P., Norman, M.E., Miller, A.D., Wang, D., et al., 2016. SIV infection-mediated changes in gastrointestinal bacterial microbiome and virome are associated with immunodeficiency and prevented by vaccination. *Cell Host Microbe* 19, 323–335.
- Handley, S.A., Thackray, L.B., Zhao, G., Presti, R., Miller, A.D., Droit, L., Abbink, P., Maxfield, L.F., Kambal, A., Duan, E., et al., 2012. Pathogenic simian immunodeficiency virus infection is associated with expansion of the enteric virome. *Cell* 151, 253–266.
- Hillman, E.T., Lu, H., Yao, T., Nakatsu, C.H., 2017. Microbial ecology along the gastrointestinal tract. *Microbes Environ.* 32, 300–313.
- Hold, G.L., Pryde, S.E., Russell, V.J., Furrer, E., Flint, H.J., 2002. Assessment of microbial diversity in human colonic samples by 16S rDNA sequence analysis. *FEMS Microbiol. Ecol.* 39, 33–39.
- Kapusinszky, B., Ardeshtir, A., Mulvaney, U., Deng, X., Delwart, E., 2017. Case-control comparison of enteric viromes in captive rhesus macaques with acute or idiopathic chronic diarrhea. *J. Virol.* 91.
- Kim, M.S., Bae, J.W., 2016. Spatial disturbances in altered mucosal and luminal gut viromes of diet-induced obese mice. *Environ. Microbiol.* 18, 1498–1510.
- Kostic, A.D., Gevers, D., Siljander, H., Vatanen, T., Hyotylainen, T., Hamalainen, A.M., Peet, A., Tillmann, V., Poho, P., Mattila, I., et al., 2015. The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. *Cell Host Microbe* 17, 260–273.
- Li, D., Chen, H., Mao, B., Yang, Q., Zhao, J., Gu, Z., Zhang, H., Chen, Y.Q., Chen, W., 2017. Microbial biogeography and core microbiota of the rat digestive tract. *Sci. Rep.* 8, 45840.
- Lim, E.S., Wang, D., Holtz, L.R., 2016. The bacterial microbiome and virome milestones of infant development. *Trends Microbiol.* 24, 801–810.
- Lim, E.S., Zhou, Y., Zhao, G., Bauer, I.K., Droit, L., Ndao, I.M., Warner, B.B., Tarr, P.I., Wang, D., Holtz, L.R., 2015. Early life dynamics of the human gut virome and bacterial microbiome in infants. *Nat. Med.*
- Looft, T., Allen, H.K., Cantarel, B.L., Levine, U.Y., Bayles, D.O., Alt, D.P., Henrissat, B., Stanton, T.B., 2014. Bacteria, phages and pigs: the effects of in-feed antibiotics on the microbiome at different gut locations. *ISME J.* 8, 1566–1576.
- Lusiak-Szelachowska, M., Weber-Dabrowska, B., Jonczyk-Matysiak, E., Wojciechowska, R., Gorski, A., 2017. Bacteriophages in the gastrointestinal tract and their implications. *Gut Pathog.* 9, 44.
- Ly, M., Abeles, S.R., Boehm, T.K., Robles-Sikisaka, R., Naidu, M., Santiago-Rodriguez, T., Pride, D.T., 2014. Altered oral viral ecology in association with periodontal disease. *MBio* 5 (e01133-0114).
- Manrique, P., Bolduc, B., Walk, S.T., van der Oost, J., de Vos, W.M., Young, M.J., 2016. Healthy human gut phageome. *Proc. Natl. Acad. Sci. USA* 113, 10400–10405.
- Manrique, P., Dills, M., Young, M.J., 2017. The human gut Phage community and its implications for health and disease. *Viruses* 9.
- Minot, S., Sinha, R., Chen, J., Li, H., Keilbaugh, S.A., Wu, G.D., Lewis, J.D., Bushman, F.D., 2011. The human gut virome: inter-individual variation and dynamic response to diet. *Genome Res.* 21, 1616–1625.
- Monaco, C.L., Gootenberg, D.B., Zhao, G., Handley, S.A., Ghebremichael, M.S., Lim, E.S., Lankowski, A., Baldrige, M.T., Wilen, C.B., Flagg, M., et al., 2016. Altered virome and bacterial microbiome in human immunodeficiency virus-associated acquired immunodeficiency syndrome. *Cell Host Microbe* 19, 311–322.
- Niu, B., Zhu, Z., Fu, L., Wu, S., Li, W., 2011. FR-HIT, a very fast program to recruit metagenomic reads to homologous reference genomes. *Bioinformatics* 27, 1704–1705.
- Norman, J.M., Handley, S.A., Baldrige, M.T., Droit, L., Liu, C.Y., Keller, B.C., Kambal, A., Monaco, C.L., Zhao, G., Fleshner, P., et al., 2015. Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell* 160, 447–460.
- Pepe, M.S., Longton, G., Anderson, G.L., Schummer, M., 2003. Selecting differentially expressed genes from microarray experiments. *Biometrics* 59, 133–142.
- Pop, M., Walker, A.W., Paulson, J., Lindsay, B., Antonio, M., Hossain, M.A., Oundo, J., Tamboura, B., Mai, V., Astrovskaya, I., et al., 2014. Diarrhea in young children from low-income countries leads to large-scale alterations in intestinal microbiota composition. *Genome Biol.* 15, R76.
- Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., et al., 2010. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464, 59–65.
- Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., Liang, S., Zhang, W., Guan, Y., Shen, D., et al., 2012. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* 490, 55–60.
- Reyes, A., Blanton, L.V., Cao, S., Zhao, G., Manary, M., Trehan, I., Smith, M.I., Wang, D., Virgin, H.W., Rohwer, F., et al., 2015. Gut DNA viromes of Malawian twins discordant for severe acute malnutrition. *Proc. Natl. Acad. Sci. USA* 112, 11941–11946.
- Reyes, A., Haynes, M., Hanson, N., Angly, F.E., Heath, A.C., Rohwer, F., Gordon, J.I., 2010. Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* 466, 334–338.
- Scalaferrì, F., Pizzoferrato, M., Pecere, S., Forte, F., Gasbarrini, A., 2012. Bacterial flora as a cause or treatment of chronic diarrhea. *Gastroenterol. Clin. N. Am.* 41, 581–602.
- Stearns, J.C., Lynch, M.D., Senadheera, D.B., Tenenbaum, H.C., Goldberg, M.B., Cvitkovich, D.G., Croitoru, K., Moreno-Hagelsieb, G., Neufeld, J.D., 2011. Bacterial biogeography of the human digestive tract. *Sci. Rep.* 1, 170.
- Virgin, H.W., Wherry, E.J., Ahmed, R., 2009. Redefining chronic viral infection. *Cell* 138, 30–50.
- Virgin, H.W., 2014. The virome in mammalian physiology and disease. *Cell* 157, 142–150.
- Wang, D., Urisman, A., Liu, Y.T., Springer, M., Ksiazek, T.G., Erdman, D.D., Mardis, E.R., Hickenbotham, M., Magrini, V., Eldred, J., et al., 2003. Viral discovery and sequence recovery using DNA microarrays. *PLoS Biol.* 1, E2.
- Yasuda, K., Oh, K., Ren, B., Tickle, T.L., Franzosa, E.A., Wachtman, L.M., Miller, A.D., Westmoreland, S.V., Mansfield, K.G., Vallender, E.J., et al., 2015. Biogeography of the intestinal mucosal and luminal microbiome in the rhesus macaque. *Cell Host Microbe* 17, 385–391.
- Yatsunencko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhin, A.P., et al., 2012. Human gut microbiome viewed across age and geography. *Nature* 486, 222–227.

- Zarate, S., Taboada, B., Yocupicio-Monroy, M., Arias, C.F., 2018. The human virome. *Arch. Med. Res.*
- Zhao, G., Vatanen, T., Droit, L., Park, A., Kostic, A.D., Poon, T.W., Vlamakis, H., Siljander, H., Harkonen, T., Hamalainen, A.M., et al., 2017a. Intestinal virome changes precede autoimmunity in type 1 diabetes-susceptible children. *Proc. Natl. Acad. Sci. USA* 114, E6166–E6175.
- Zhao, G., Wu, G., Lim, E.S., Droit, L., Krishnamurthy, S., Barouch, D.H., Virgin, H.W., Wang, D., 2017b. VirusSeeker, a computational pipeline for virus discovery and virome composition analysis. *Virology* 503, 21–30.
- Zoetendal, E.G., Raes, J., van den Bogert, B., Arumugam, M., Booijink, C.C., Troost, F.J., Bork, P., Wels, M., de Vos, W.M., Kleerebezem, M., 2012. The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates. *ISME J.* 6, 1415–1426.