6-28-2018

Diversity of ssDNA Phages Related to the Family *Microviridae* within the *Ciona robusta* Gut

Alexandria Creasy
University of South Florida, creasy@mail.usf.edu

Diversity of ssDNA Phages Related to the Family *Microviridae* within the *Ciona robusta* Gut

by

Alexandria Creasy

A thesis submitted in partial fulfillment
of the requirements of the degree of
Masters of Science in Marine Science
College of Marine Science
Biological Oceanography
University of South Florida

Major Professor: Mya Breitbart, PhD
Co advisor: Larry J Dishaw, PhD
Karyna Rosario, PhD

Date of Approval:
June 16, 2018

Keywords: *Gokushovirinae*, gut model, virome, tunicates, filter feeder

Table of Contents

List of figures

Abstract

The gut microbiome is a complex ecosystem of bacteria, viruses, and fungi that strongly influences animal health. The bacterial component, for example, contributes orders of magnitude more gene products to host physiology than the host genome; thus, changes to the composition of these bacterial communities can have profound influences on the health of the animal. By infecting and lysing their hosts, viruses (particularly viruses infecting bacteria or phages) can affect critical functions in these environments, yet the consequences of these infections remain to be fully described. Most studies investigating gut viromes to date have focused on double-stranded DNA (dsDNA) phages and, consequently, little is known about the smaller single-stranded DNA (ssDNA) phages, which also inhabit gut environments. In this study, we investigated ssDNA phages of the *Microviridae* family within the gut of an invertebrate organism, *Ciona robusta*, used as a model system to better understand gut microbial interactions. As a filter feeder, *Ciona* concentrates dissolved organics and microbes as part of its diet, yet maintains a microbiome distinct from the surrounding water column. We identified 258 unique ssDNA phage genomes representing a diversity of *Microviridae* subgroups including novel members of the established *Gokushovirinae* subfamily and several proposed phylogenetic groups (*Alpavirinae, Aravirinae,* Group D, Parabacteroides prophages, and *Pequeñovirus*). Over 70% of the genomes belonged to the *Gokushovirinae*; however, 155 of these genomes did not group with previously described sequences. Our results highlight an unprecedented diversity of ssDNA phages from an animal gut. Furthermore, comparative analysis between samples collected from *Ciona* specimens with full and cleared guts as well as the surrounding water indicated that *Ciona*

retains a unique and highly diverse community of ssDNA phages. The present study significantly expands the known diversity within the *Microviridae* family and suggests that *Ciona* is a promising system for studying the role of ssDNA phages within animal guts.

Introduction

Recent studies of host-microbe interactions have recognized the importance of the holobiont, which acknowledges the complex partnerships between an animal and the entirety of its associated microbial communities (Theis *et al.* 2016). These host-microbial network-type interactions form a complex dynamic relationship impacting host physiology, including the procurement of nutrients and metabolic output (Nicholson *et al.* 2012, Bordenstein and Theis 2015). One such complex, bacteria-rich, ecosystem is the nutrient- and mucus-rich gut of animals where the microbiome is known to contribute orders of magnitude more gene products to host physiology than the host genome itself (Turnbaugh *et al.* 2007). Although many of the metabolic contributions are predicted from the cellular component of the microbiome (i.e., bacteria, fungi and archaea), it is now recognized that viruses also play an important role in microbiome dynamics.

Viruses present in gut communities are dominated by bacteriophages (phages), which are viruses that infect bacteria (Breitbart *et al.* 2003, Reyes *et al.* 2010, Minot *et al.* 2011, Manrique *et al.* 2016, Manrique *et al.* 2017, Mirzaei and Maurice 2017, Yutin *et al.* 2018). Phages are thought to be in a 1:1 ratio with bacteria in the human intestine (Reyes *et al.* 2010, Kim *et al.* 2011, Minot *et al.* 2011), which contrasts with ratios observed in environmental samples (e.g., 10:1 in sea water (Suttle 2007)). Since phages can have dramatic influences on the structure and function of microbial communities, through lysis of bacteria and horizontal gene transfer (Suttle 2007, Shapiro *et al.* 2010, Koskella and Brockhurst 2014), , phage-bacterial interactions appear crucial in all environments studied to date (Rohwer and Thurber 2009, Shapiro *et al.* 2010,

Breitbart 2011, Scarpellini *et al.* 2015, Thurber *et al.* 2017), including microbial communities that are found in close association with animals (Turnbaugh *et al.* 2007, Dishaw *et al.* 2014 Isaacson, 2012 #335, Sanders *et al.* 2015, Soverini *et al.* 2016). While phage dynamics likely have dramatic influences on the physiology of the animal host, very little is known about the role of viromes (the cumulative viral community associated with a given host) or their impact within the gut environment (Reyes *et al.* 2010, Mills *et al.* 2013, Lim *et al.* 2015, Ogilvie and Jones 2015, Carding *et al.* 2017, Mirzaei and Maurice 2017).

Although a variety of factors may influence the abundance, diversity, and ecological roles of phages in a complex microcosm like the gut (Yatsunenko *et al.* 2012, O'Toole and Jeffery 2015), the complete diversity of phages in the gut is difficult to determine due to a lack of universal gene markers within their genomes (Rohwer and Thurber 2009). Viral metagenomics, where the collective viral nucleic acids from a given sample are sequenced, is an efficient alternative approach to exploring the phage fraction of the gut virome (Breitbart *et al.* 2003, Kim *et al.* 2011). Although both single- (ss) and double-stranded (ds) DNA phages have been identified in animal guts (Lim *et al.* 2015, Reyes *et al.* 2015), the diversity of dsDNA phages remains the best characterized due to methodological biases towards them. Early viromic studies used linker-amplified shotgun sequencing approaches that biased against ssDNA viruses (Roux *et al.* 2016); however, the implementation of rolling circle amplification (RCA) to obtain enough DNA for sequencing has revealed a large diversity of genomes from ssDNA viruses in the guts of humans and other animals (Reyes *et al.* 2010, Kim *et al.* 2011, Krupovic and Forterre 2011, Minot *et al.* 2011, Roux *et al.* 2012, Reyes *et al.* 2013, Waller *et al.* 2014, Reyes *et al.* 2015, Guo *et al.* 2017, Moreno *et al.* 2017, Tikhe and Husseneder 2017, D'arc *et al.* 2018).

Studies investigating microbial communities in the human gut have shown that the microbiome structure seen in healthy newborns develops and changes within the first 2-3 years of life from a nearly sterile gut environment to a dynamic community maintained throughout adulthood (Lim *et al.* 2015). Human gut communities are dominated by the dsDNA phage from the order *Caudovirales* and the ssDNA phage family *Microviridae* (Breitbart *et al.* 2008, Minot *et al.* 2013, Lim *et al.* 2015). A recent study looking into the early life virome in infants found the phage richness and diversity is greatest at 0 months and proceeds to decrease with age (Lim *et al.* 2015). Around 24 months of age the phage community shifts to an increased richness in *Microviridae* showing early infant development is marked by a decrease in overall phage richness and diversity along with a shift in phage community towards predominately *Microviridae*. These gut communities can be affected through diseased states, such as inflammatory bowel disease, and can be correlated with a decrease in the *Microviridae:Caudovirales* ratio, with an increase towards *Caudovirales* (Norman *et al.* 2015). These recent studies have highlighted *Microviridae* phages as important members within the gut system that may be associated with an individual's health state (Kim *et al.* 2011, Roux *et al.* 2012, Lim *et al.* 2015, Norman *et al.* 2015, Reyes *et al.* 2015). However, little is known about the potential roles of members of the *Microviridae* in the gut microbiome.

Studies leveraging simpler model systems to characterize the virome can help lead to hypothesis-driven experimental approaches to dissect these multifaceted biological and ecological processes. We have been developing a marine cosmopolitan sea squirt species, *Ciona robusta* (formerly *Ciona intestinalis* subtype A), in efforts to interrogate gut microbiome dynamics. This sessile, invertebrate chordate is a well-studied developmental model (Satoh *et al.* 2003) with a sequenced genome (Dehal *et al.* 2002). Because *Ciona* is a filter feeding organism,

its gut is a microcosm of microbial interactions that experience vast and continuous exposures to the large microbial and viral diversity found in seawater. Previous efforts identified remarkable stability among some elements of the microbiome, which are distinct from surrounding seawater (Dishaw *et al.* 2014, Leigh *et al.* 2018). *Ciona* has structurally distinct gut compartments (stomach, midgut, hindgut) and maintains core bacterial species, some of which exhibit compartmentalization (Dishaw *et al.* 2014). A recent effort characterizing the *Ciona* gut virome revealed a viral community encompassing 23 different families within the Order *Caudovirales*. There was also evidence that some of the viruses (with genomes >5 kilobases (kb)) were restricted to some gut compartments, an observation also noted among some bacterial communities (Leigh *et al.* 2018). Together, these findings suggest that strong selective pressures operate within the gut of this simple model organism. However, additional viral groups such as ssDNA viruses have yet to be characterized within *Ciona,* a finding that would support the use of this model system in studies to investigate the role of ssDNA phages in animal guts.

The main objectives of the current study were to: 1) characterize the diversity of complete circular genomes related to the *Microviridae*, the most widely detected ssDNA phages among animals, from the *Ciona* gut, 2) evaluate if identified ssDNA phages were unique to *Ciona* or if they could also be identified in water samples, and 3) evaluate if there were unique ssDNA phage assemblages in the different gut compartments. We show that *Ciona* harbors a diverse community of small ssDNA phages that is distinct from the water column and report novel viral genomes that significantly expand the known diversity of the *Microviridae* family.

Methods

Sample collections and library preparations

Sequences described in this manuscript are derived from viromes generated by Leigh *et.al.* (2018), which focused on analysis of dsDNA diversity. Animal guts were sampled from *Ciona robusta* harvested near San Diego, CA (M-Rep, Carlsbad, CA, USA). Ten animals were selected at random; five specimens were placed into virus-free 100kD-filtered artificial seawater (Grzenia *et al.* 2008) to clear guts of dietary contents (water changed every 4 hours for 24 hours). These samples are referred to as 'cleared guts.' The remaining five animals were dissected with full gut contents. All animal guts, full (F) or cleared (C), were tri-sected (stomach (S), midgut (M), hindgut (H)) and snap-frozen in liquid nitrogen. Collected tissues from each gut type (n = 5) were disrupted in 3 mL of sterile suspension buffer using the GentleMACS dissociator (Miltenyi Biotec, Bergisch Gladbach, Germany). Tissue fragments were pelleted (6,000 xg for 10 minutes) and the supernatant consisting of the cellular fraction was filtered a 0.22 μm Sterivex filter (Merck Millipore) and the filtrate containing virus-like particles (VLPs) was collected. In addition to the viral fraction from *Ciona* guts, viruses from surrounding seawater in Mission Bay (MB), as well as the flow-through holding-tank water (CB), were processed to determine if viruses detected in *Ciona* guts could be detected in the surrounding water. For this purpose, one liter of seawater was filtered through a 0.22 μm Sterivex filter and VLPs in the filtrate were concentrated to 1 mL using a 100 kDa Amicon Centrifugal filter (EMD, Merck Millipore). VLPs were purified and further concentrated via cesium chloride (CsCl) gradient centrifugation (Thurber *et al.* 2009) and collecting the 1.2-1.5 g/mL fraction with sterile syringe and needle into

a 2 mL sterile tube. To remove potential bacteria or vesicles still present in the sample, chloroform (final concentration 20% vol/vol) was added to the viral fraction and incubated at room temperature for 10 minutes. Samples were then centrifuged for 30 seconds at maximum speed (20,000 xg) and the top aqueous layer was recovered. Unencapsidated, free nucleic acids were then removed by treating with DNase I (2.5 U/µL final concentration) for 3 h at 37°C with frequent vortexing; the nuclease was inactivated by treating with 0.5M EDTA pH 8.0 (final concentration 20 mM). Purified VLP samples were tested to rule out bacterial contamination by PCR amplification of the 16S rRNA gene using primers 27F and 1492R (Weisburg *et al.* 1991) and epifluorescence microscopy. Viral DNA was then extracted from 200 µl of the viral concentrate using the Qiagen MinElute Virus Spin Kit (Qiagen, Inc., Valencia, CA, USA) and amplified via rolling circle amplification (RCA) using the GenomiPhi V2 DNA Amplification Kit (GE Healthcare Life Sciences, Pittsburgh, PA, USA), resulting in ~1 µg viral DNA per sample. Three identical RCA reactions per sample were prepared and pooled for sequencing. Qubit (Thermo Fisher Scientific, Waltham, MA, USA) was used to obtain viral DNA concentration and amplification was verified with 1% agarose gel electrophoresis. Final, amplified products were cleaned via MinElute PCR Purification Kit (Qiagen, Inc.). DNA quality and quantity were assessed using the BioAnalyzer 2100 (Agilent Technologies, Santa Clara, CA). Sequencing was performed on an Illumina MiSeq platform generating mate-pair (2x250 bp) libraries (Operon, Eurofins MWG Operon LLC, Huntsville, AL). Sequences were analyzed in the CyVerse Cyberinfrastructure using different bioinformatics applications (Apps) (Goff *et al.* 2011). Briefly, raw sequences were trimmed based on quality scores using the Trimmomatic App version 0.35.0 (Bolger *et al.* 2014) and quality-filtered sequences were then assembled using the SPAdes App version 3.6.0 (Bankevich *et al.* 2012). Contigs were screened with the

VirSorter App to detect potential viral sequences. VirSorter outputs were uploaded to MetaVir (Roux *et al.* 2014) (IDs 7811 (SF), 8143 (SC), 7815 (MF), 7814 (MC), 7812 (HF), 7910 (HC), 7816 (CB), 7819 (MB)) (Leigh, *et.al.* 2018).

Analysis of *Microviridae*

Identification and annotation of *Microviridae* genomes

Since members of the *Microviridae* have circular genomes, circular contig sequences were identified using MetaVir (Roux *et al.* 2011). Circular contig sequences ranging from 1 kb to 8 kb in length were compared (BLASTx, e-value < 0.0001) against a curated database of 4120 *Microviridae* major capsid protein (MCP) amino acid sequences (Vincent *et al.*, unpublished). Contig sequences with significant matches in the *Microviridae* MCP database were then compared (BLASTx, e-value < 0.001) against the Genbank non-redundant (nr) database to eliminate contig sequences that had better matches to cellular organisms (i.e., false positives). BLASTx outputs were explored using the MEGAN community edition software v6.8.9 (Huson *et al.* 2016) to identify sequences related to the *Microviridae*. *Microviridae*-related sequences were then trimmed to unit length genomes manually by identifying repeated sequences. Unit-length genomes were annotated using Geneious v10.1.3 (Kearse *et al.* 2012). For this purpose, open reading frames (ORFs) encoding putative proteins >80 amino acids (aa) were compared against the Genbank nr database using BLASTp (e-value <0.001). All genomes were manually edited to start at the start codon of the MCP in subsequent alignments. Genome-wide and MCP pairwise identities were calculated using the sequence demarcation tool (SDT) v1.2 (Muhire *et al.* 2014). Genomes and MCP were dereplicated at 95% nucleic acid and amino acid identity, respectively.

Genome comparisons and phylogenetic analysis

Reference MCP amino acid sequences, including VP1 (*Gokushovirinae*) and Protein F (*Bullavirinae*), were collected from GenBank. These reference sequences also contained select *Microviridae*-like sequences identified from metagenomes (Rosario *et al.* 2012, Roux *et al.* 2012, Roux *et al.* 2012, Labonté and Suttle 2013, Bryson *et al.* 2015, Quaiser *et al.* 2015, Walters *et al.* 2017) and those integrated into bacterial genomes (Krupovic and Forterre 2011) (see Supplementary Table 1 for details). Reference MCP sequences were aligned with sequences identified in *Ciona* guts using MUSCLE (Edgar 2004) as implemented in Geneious v10.1.3 (Kearse *et al.* 2012) and manually edited. A maximum likelihood tree was then created using PhyML with aLRT-like probabilities for branch support (Lefort *et al.* 2017) and visualized with FigTree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/). Branches with probability values less than 0.70 were collapsed via TreeGraph 2 (Stöver and Müller 2010).

To evaluate the diversity of *Microviridae* in different *Ciona* gut compartments, a recruitment analysis was performed following the pipeline suggested for analysis of viral abundance and distribution through iVirus as implemented in the CyVerse Cyberinfrastructure (Bolduc *et al.* 2017). For this purpose, The Bowtiebatch v1.0.0 and Read2Ref v1.0.1 Apps were used with default parameters to map overlapping genomes between the gut compartments (Bolduc *et al.* 2017). The output table displaying relative abundances was converted to a binary matrix and used to assess the number of shared *Microviridae* genomes in the different compartments based on presence and absence. This information was summarized using the Venn Diagram package (Chen and Boutros 2011) with community indexes calculated using a binary matrix and visualized as a dendrogram using the Vegan package (Oksanen *et al.* 2007) created in R v3.3.2 (Team 2014). The binary data was analyzed with three methods (Bray-Curtis index,

Jaccard index, Euclidian dissimilarity), each resulting in congruent tree topologies. Note that the number of reads recruiting to a given genome was not considered either in this study or in comparing these genomes to the dsDNA phages (Leigh *et al.* 2018), since there are known biases created by RCA that lead to overrepresentation of ssDNA circular genomes (Kim *et al.* 2008, Kim and Bae 2011, Roux *et al.* 2016).

Results

Diversity of *Microviridae*-like genomes in *Ciona* gut compartments

Analysis of the *Ciona* gut viromes, which included six libraries representing viral sequences from three gut compartments (stomach, midgut, and hindgut) from cleared and full guts, revealed 488 circular contig sequences (1 – 8 kb in length) with BLAST similarity to *Microviridae*-related genomes. A total of 258 unique genomes were identified after de-replicating these sequences based on 95% genome-wide pairwise identities. Unique genomes were then named with the prefix, *Ciona* gut microphage (CGM), followed by a number (Supplementary Table 2). The average size of the identified CGM genomes was ~ 4.3 kb with a range from 3.9 kb to 5.8 kb, which is consistent with previously described members of the *Microviridae* family (Doore and Fane 2016).

To assess diversity, each CGM genome was annotated and MCP amino acid sequences were used for phylogenetic analysis (Fig. 1). Based on this analysis, the vast majority (n=188) of CGM sequences grouped with the established subfamily *Gokushovirinae* (Lefkowitz *et al.* 2017), followed by sequences closely related to the proposed Group D microviruses (n=33) (Quaiser *et al.* 2015) and the subfamily *Pichovirinae* (n=19) (Roux *et al.* 2012). A smaller proportion of

CGM sequences were related to Parabacteroides prophages (n=7) (Sakamoto and Benno 2006, Quaiser *et al.* 2015), and members of the proposed *Aravirinae* (n=1) (Quaiser *et al.* 2015) and *Alpavirinae* (n=8) subfamilies (Krupovic and Forterre 2011, Roux *et al.* 2012). Six of the CGM MCP sequences (CGM_251, CGM_223, CGM_252, CGM_222, CGM_249, CGM_250) within the *Alpavirinae* are grouping paraphyletically and seem distinct from known sequences from this subgroup. No CGM MCPs were found to group with the *Stokavirinae* (Quaiser et al. 2015) or *Bullavirinae* (Lefkowitz *et al.* 2017) subgroups. However, two CGM sequences were most

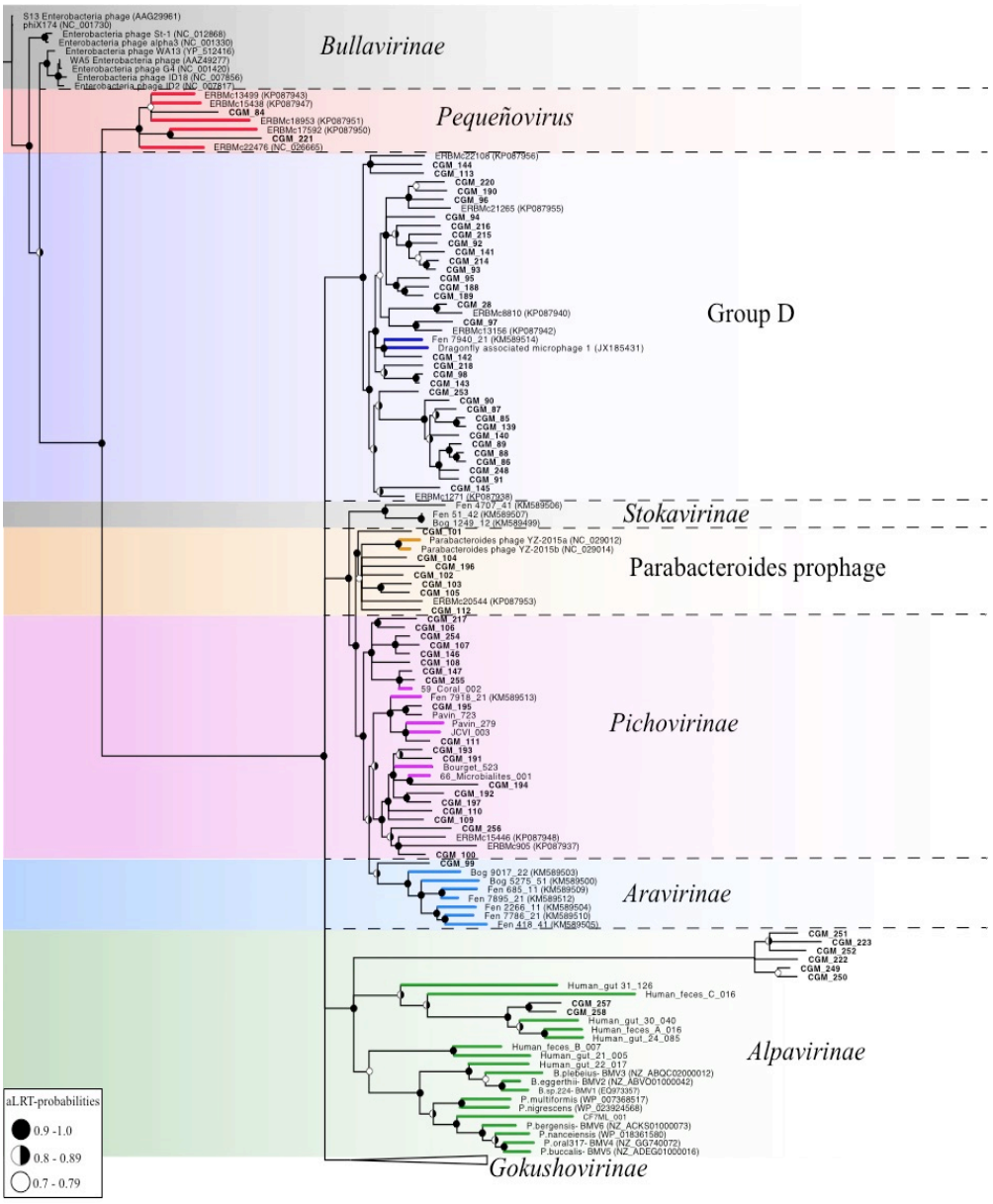

Figure 1. Maximum likelihood phylogenetic tree of predicted major capsid protein (MCP) sequences from the *Ciona* gut *Microviridae* (CGM, n = 258) along with representative sequences from previously described (proposed) subfamilies (n = 96). The tree was created using PhyML with aLRT-probabilities. Branches with probability values less than 0.7 were collapsed. Values greater than 0.7 are indicated at nodes. Suggested subfamily demarcations are delineated with dashed lines and colors based on previously classified sequences. Subfamilies for which CGM sequences were not identified are highlighted in grey color. Note: *Gokushovirinae* sub-tree is displayed in Figure 2.

closely related to a sister clade of the *Bullavirinae* subfamily, the *Pequeñovirus* group (Bryson *et al.* 2015).

Since 73% of the CGM genomes group within the *Gokushovirinae* subfamily, this subfamily is presented in a separate tree (Fig. 2). The majority (82%, n=155) of the CGM MCP sequences group within the *Gokushovirinae* subfamily do not associate with previously reported sequences and share less than 70% aa identify with known *Gokushovirinae*. The gene synteny among reported *Microviridae* and CGM genome sequences were compared to evaluate if CGM genomes possess novel genome organizations (Fig. 3). The CGM genomes expand upon previously identified gene synteny, yet many sequences share syntenic similarity with representative or published sequences. The most diverse group, in terms of genome organization, was Group D with 10 different gene synteny patterns.

Structure of *Microviridae*-like communities

Richness of *Microviridae*-like genomes within gut compartments was compared among animals with either full or cleared guts. Stomach clear (SC) and midgut clear (MC) contained the highest number of unique genomes within the gut; these two compartments also share the largest number of genomes (Fig. 4). Both water samples (MB & CB) group separately from the gut compartments (Fig. 4), but all CGM sequences found in the water samples are also found throughout the gut compartments and all belong to the *Gokushovirinae* subgroup (Supplementary Table 1). Both the full and clear hindgut samples share similarity with the midgut full (MF); however, the HF and MF share 88 genomes, the highest degree of overlap between the full gut compartments. The SC contained 211 *Microviridae* genomes; the largest number seen in any cleared gut compartment, while the HC had the lowest number of genomes (n=123) of the

cleared compartments, only 8 of which were unique. Interestingly, despite being full of dietary material, the full gut compartments have a lower overall richness than the cleared. Four of the sequences that fall within the *Alpavirinae* subfamily (CGM_251, CGM_223, CGM_250, CGM_257) were only found in the hindgut full (Supplementary Table 2), and not seen in any other gut compartments. MF has the highest richness within the full gut, with a total of 132 genomes and 23 unique to that compartment. The lowest richness among the full compartments was found in the stomach, with 107 genomes and only 19 unique. For details on compartmentalization, see Supplementary Table 2.

Discussion

The gut microbiome of animals is rich in diverse microbial communities (Turnbaugh *et al.* 2007, Dishaw *et al.* 2014). Most microbiome research has focused on the bacterial communities, with descriptions of the virome now gaining traction (Reyes *et al.* 2012, Lim *et al.* 2015, Ogilvie and Jones 2015). Understanding the gut virome is relevant to both the host and the cellular microbiome because viruses, whether infecting eukaryotic, bacterial or archaeal hosts, can have profound influences in shaping gut homeostasis (Scarpellini *et al.* 2015, Carding *et al.* 2017, Dahiya 2017). Here, we described ssDNA phages found in the *Ciona* gut in an effort to further characterize the virome of this invertebrate model organism. Leveraging recent applications of

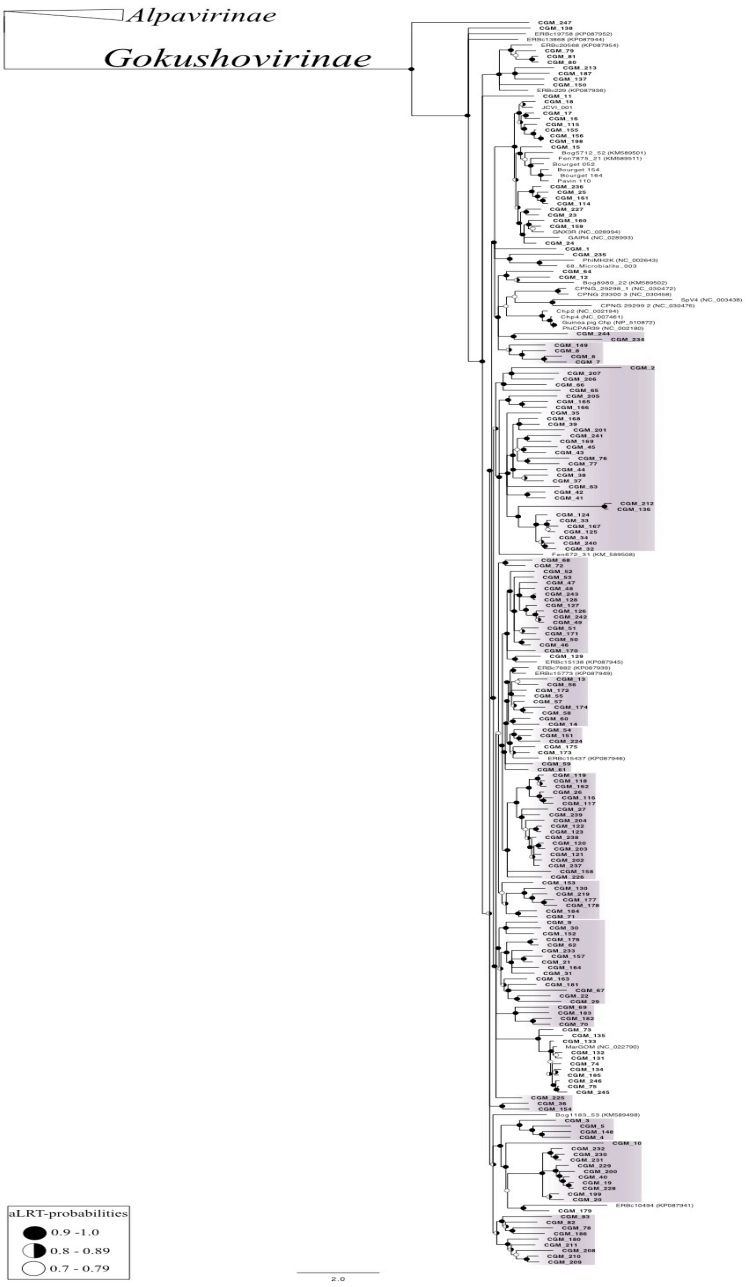

Figure 2. Maximum likelihood phylogenetic tree of predicted major capsid protein (MCP) sequences from the *Ciona* gut *Microviridae* (CGM, n = 188) that clustered within the established *Gokushovirinae* subfamily. MCP sequences representing Alpavirinae were used as an outgroup. The tree was created via PhyML with aLRT-probabilities. Branches with probability values less than 0.7 were collapsed. Values greater than 0.7 are indicated at nodes. Clades highlighted in grey represent those where CGM sequences do not group with any previous described MCP sequences.

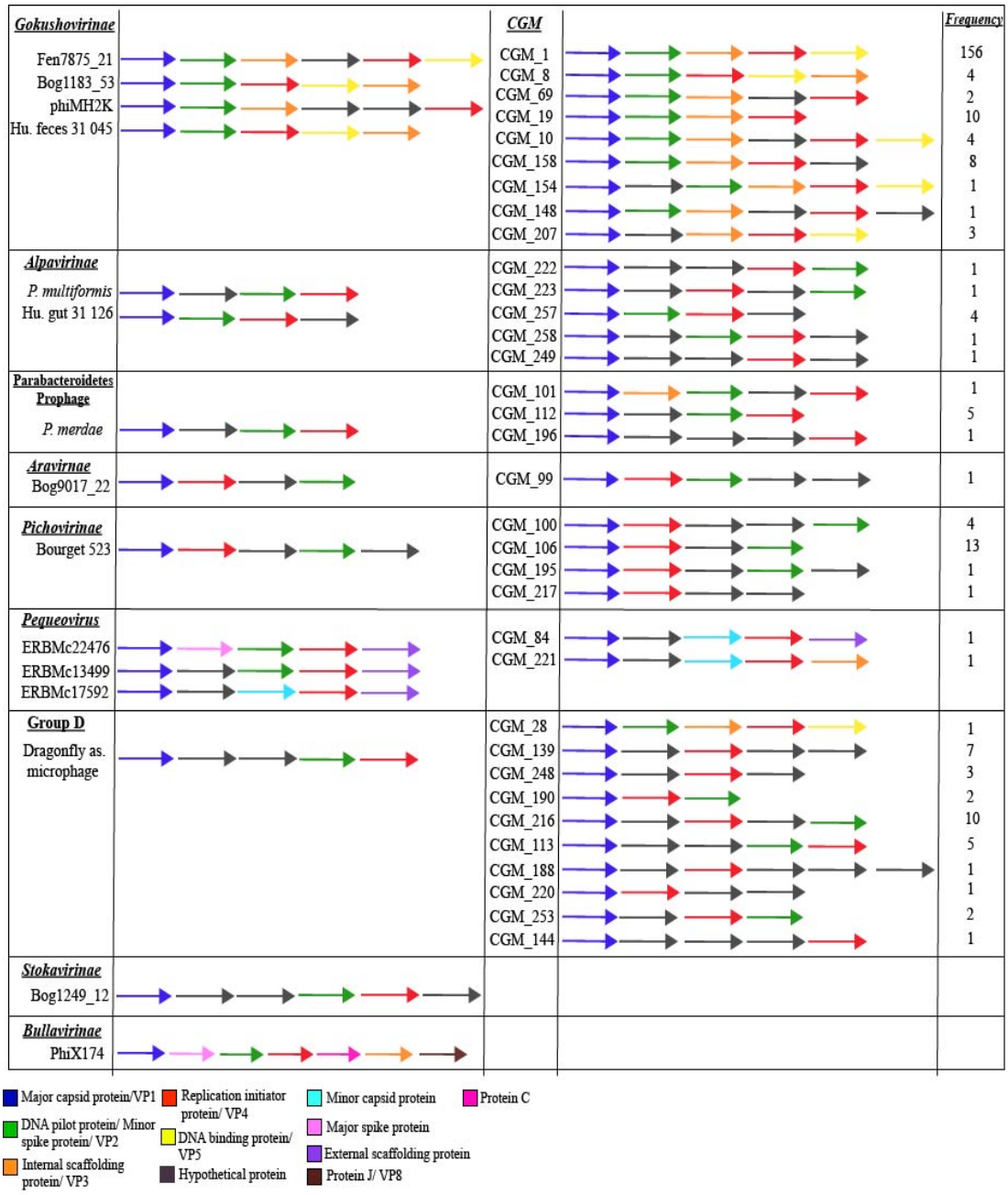

Figure 3. Gene synteny comparisons between previously described *Microviridae* genomes (left) and those discovered in the *Ciona* gut (right; CGM). All genomes were manually annotated to start at the major capsid protein (MCP) and open reading frames (ORFs) > 80 aa are shown in linear fashion (i.e., overlapping genes are shown in order based on the position of the start codon). ORFs are color-coded based on PHA (phage protein subset of the Entrez protein cluster (PRK) database. One representative of each gene order known to exist within a given (proposed) subfamily is shown, and the numbers of CGM genomes containing a particular gene order are specified on the far right.

RCA in virome studies has dramatically increased our discovery of smaller, ssDNA viruses including the *Microviridae* phages. These viruses have now been described in a variety of diverse habitats, including animal guts (Jørgensen *et al.* 2014, Moreno *et al.* 2017, Tikhe and Husseneder 2017, Walters *et al.* 2017, D'arc *et al.* 2018), human guts (Zhang *et al.* 2005, Breitbart *et al.* 2008, Reyes *et al.* 2010, Kim *et al.* 2011, Minot *et al.* 2011, Reyes *et al.* 2013, Waller *et al.* 2014, Lim *et al.* 2015, Reyes *et al.* 2015, Santiago-Rodriguez *et al.* 2015, Guo *et al.* 2017, McCann *et al.* 2018), reclaimed water (Rosario *et al.* 2009), sewage (Hopkins *et al.* 2014, Pearson *et al.* 2016), fresh water systems (Kim *et al.* 2008, López-Bueno *et al.* 2009, Roux *et al.* 2012, Hopkins *et al.* 2014, Zhong *et al.* 2015), marine systems (Breitbart *et al.* 2002, Angly *et al.* 2006, Bench *et al.* 2007, Tucker *et al.* 2011, Labonté and Suttle 2013, Labonté *et al.* 2015, Yoshida *et al.* 2018, Yu *et al.* 2018), methane seeps (Bryson *et al.* 2015), modern stromatolites (Desnues *et al.* 2008), confined aquifers (Smith *et al.* 2013), sediments (Kim *et al.* 2008, Yoshida *et al.* 2013, Reavy *et al.* 2015, Han *et al.* 2017, Yoshida *et al.* 2018) dragonflies (Rosario *et al.* 2012), and fruit trees (Basso *et al.* 2015). Despite this rapid increase in sequence information for ssDNA viruses, their identification in such diverse environments has yet to reveal information about their functions in these systems. Now that a diversity of *Microviridae* phages has been identified in *Ciona*, future experiments using this model organism will aim to define the role of these viruses in an animal gut.

In characterizing the diversity of *Microviridae*-like contigs, we found 258 unique viral genomes within the *Ciona* gut, expanding the total diversity of *Microviridae* described for any single animal gut study (Reyes *et al.* 2010, Roux *et al.* 2012, Waller *et al.* 2014, Reyes *et al.* 2015, Guo *et al.* 2017, Tikhe and Husseneder 2017, D'arc *et al.* 2018, McCann *et al.* 2018). Although these studies identified *Microviridae*- like sequences, the diversity of these sequences

is often overlooked. When diversity is assessed, *Gokushovirinae* is the most readily identified subfamily. Based on the MCP phylogenetic analysis, which is typically used as a phylogentic marker for this viral group (Desnues *et al.* 2008, Roux *et al.* 2012, Labonté and Suttle 2013, Hopkins *et al.* 2014), we found a variety of novel *Microviridae* groups. The CGM diversity encompassed the recognized subfamily *Gokushovirinae* and 6 of the 9 proposed subfamilies in the literature, with *Pichovirinae,* Group D, Parabacteroides having the largest expansion.

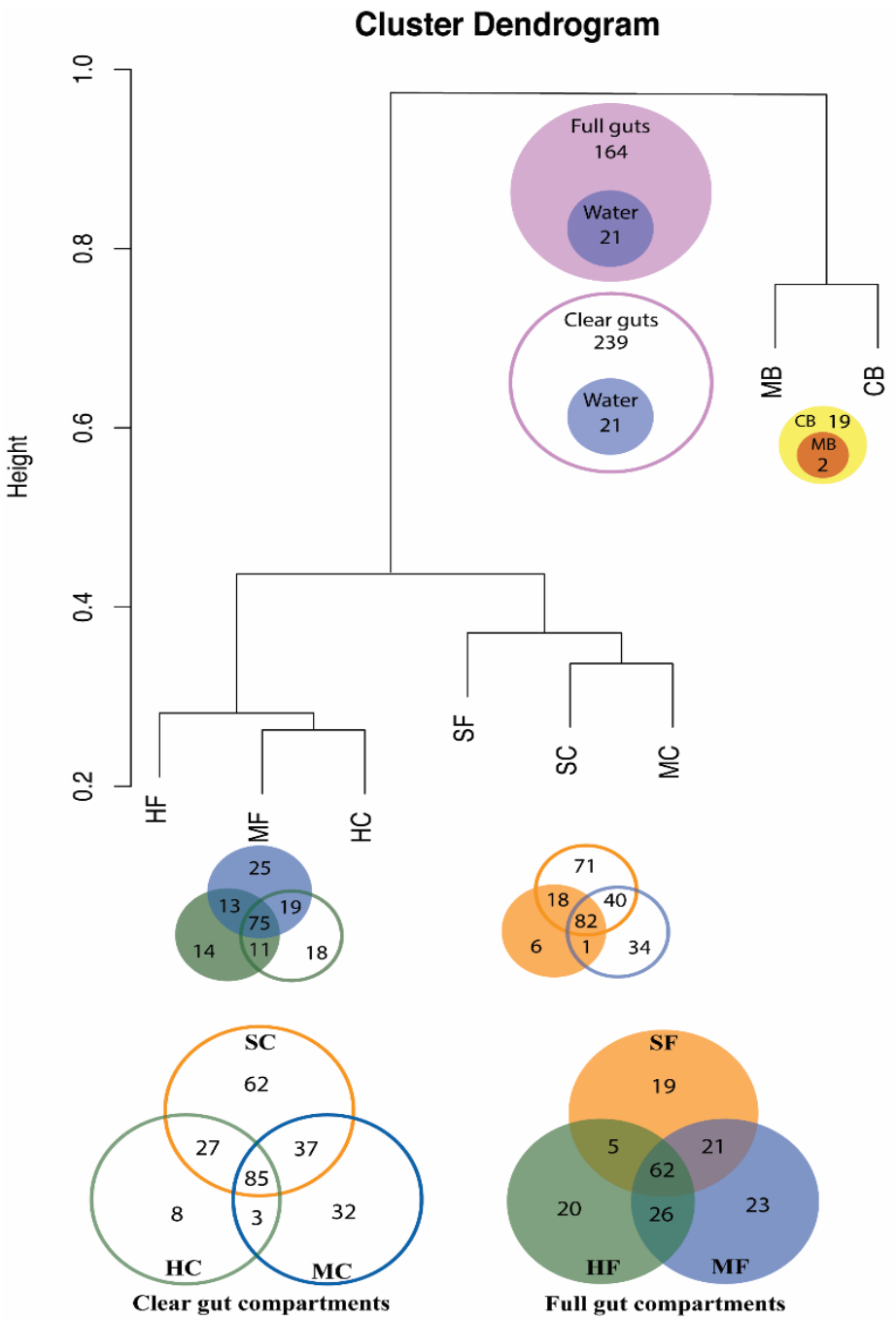

Figure 4. Cluster dendogram showing the relatedness among the CGM phage communities in the *Ciona* gut compartments and the surrounding water. The three-way Venn diagrams specify shared and unique genomes detected in each of the compared groups. The diagrams were created based on the presence/absence of CGM genomes alone. The dendogram was created using the Bray-Curtis dissimilarity index and the scale bar represents the dissimilarity values.

Interestingly, out of the 188 CGM *Gokushovirinae* genomes, 82% of those form their own clades (Fig. 2), without any representatives from other systems examined to date. Recent studies have discovered novel *Gokushovirinae* diversity (Kim *et al.* 2011, Roux *et al.* 2012, Waller *et al.* 2014, Bryson *et al.* 2015, Labonté *et al.* 2015, Reyes *et al.* 2015, Zhong *et al.* 2015, Guo *et al.* 2017, Yu *et al.* 2018), with CGM genomes being no exception. This subfamily appears predominant throughout many environments. *Gokushovirinae* are only known to infect obligate intercellular bacteria (Cherwa and Fane 2001) and have been isolated from *Bdellovibrio (Bdellomicrovirus), Chlamydia (Chlamydiamicrovirus)* and *Spiroplasmas (Spiromicrovirus)* (Doore and Fane 2016). The 16S ribosomal RNA gene data from a previous study done on the *Ciona* gut (Dishaw *et al.* 2014) found representatives from each of the families of bacteria gokushoviruses are known to infect. A limited known-host range (i.e., intracellular parasitic bacteria) paired with these findings, suggests that *Ciona* could be concentrating hosts for members of the *Gokushovirinae*, providing a model for further studies into these dynamics.

Many of the CGMs were found to share a significant degree of synteny, containing previously recognized genomic arrangements (Fig. 3). However, unique genome organizations were observed for several of the CGMs. The largest number of unique genome organizations was noted for *Gokushovirinae* and Group D viruses. In general, the observed genomic features are consistent with what is known from other previously characterized *Microviridae*. *Bullavirinae* and *Gokushovirinae* subfamilies, which are clearly distinguished by their MCP and scaffolding proteins. While most small viruses (T≤3) do not encode scaffolding proteins, larger viruses generally utilize at least one. *Bullavirinae* uniquely utilizes 2 scaffolding proteins, internal scaffolding protein (protein B) and a external scaffolding protein (protein D). The external scaffolding protein lattice is comprised of 4 protein D molecules and are in contact with 60 major

capsid proteins (protein F/VP1), along with the major spike proteins (protein G) interacting within an asymmetric arrangement. These interactions keep the procapsid together (Doore and Fane 2016). *Gokushovirinae* genomes do not encode an external scaffolding (protein D) or a major spike protein (protein G). Based on this, the vast majority of the CGM genomes belong to the *Gokushovirinae* subfamily. Noted exceptions are with the CGM genomes grouping with the *Bullavirinae* sister clade, *Pequeñovirus*, which have an external scaffolding protein. This feature is not seen among other CGM genomes or subfamiles and correlates with the known syteny from the *Pequeñovirus* genomes (Bryson *et al.* 2015).

Surprisingly, the large diversity of CGMs was mostly absent from the surrounding seawater. The MB and CB water samples only contain 2 and 21 unique ssDNA phage genomes, respectively, all of which were present in the *Ciona* gut. The CB water originates from the holding tanks the animals are placed in after field collections, before shipping. Though the animals spend less than 8 hours in these waters, they are passing water through their siphons and feeding and releasing stool pellets, which could potentially contribute to the increased number of CGM genomes seen in this sample. While it is possible that some *Microviridae* virions are too small to be captured (and concentrated) on a 100kDa filter, the more likely explanation is that these viruses are less prevalent in seawater than the *Ciona* guts. While the animals were harvested from the MB waters, nearly all the viruses (235 viral genomes) were unique to the *Ciona* gut (Fig. 4). The previously described dsDNA phages from the *Ciona* gut were also significantly different from those in the water column (Leigh *et al.* 2018). No significant correlation was noted between taxonomic classification of *Microviridae* and gut compartmentalization in this study; diverse environmental factors may influence the structure of these systems but none appear to influence how these viruses are dispersed. However, distinct

viral signatures are still found among the stomach, midgut and hindgut compartments that can inform us of similarity among these unique niches and may provide clues as to how some of these specific viruses and/or their hosts are distributed. For example, while a large number of viruses are predominately found in the stomach, the midgut clear compartment is more closely related to the stomach (clear and full) while the midgut full more closely resembles the hindgut (clear of full) (Fig. 4). These findings suggest that the midgut is likely an intermediate reservoir of viruses and that some level of compartmentalization exists among a portion of the viral communities.

Clearing of animals is a process used to void the gut of dietary and fecal material, but this process is inherently stressful to a filter feeder because food is restricted from their diet. This stress could, in part, account for the higher number of diverse viral types recovered from the SC, suggesting that the process of clearing liberates viruses from the mucosal lining of the gut that otherwise would be under-sampled when the gut is full of dietary material. Retention seems to vary from the stomach to the hindgut as the numbers of unique *Microviridae* genomes dramatically diminishes as one reaches the most distal areas of the gut. This trend is not seen within the full compartments, where the rapid transit of dietary and fecal material through the gut likely impacts the distribution and/or compartmentalization of some viruses; e.g., laboratory feeding experiments performed with fluorescently tagged food particles and/or bacteria reveals food pellets exiting the animals within 45min-1hr (Dishaw *et al.* unpublished). This rapid transit is hypothesized to impact the stability of some of these niches and likely diminishes compartmentalization of viral communities. The stress of clearing could also induce prophages. It was hypothesized previously that an increased prevalence of temperate phages may be due to prophage induction caused by the stress of clearing (Leigh *et al.* 2018). Although originally

thought to be strictly lytic, the discovery of *Microviridae* with the likelihood of integrating into bacterial genomes was noted in an *in silico* study that identified MCP-coding genes and other sequences related to the subfamily, *Gokushovirinae,* within bacterial genome sequences from various *Bacteroidetes* species common in the human mouth and gut (Krupovic and Forterre 2011). These *Bacteroidetes* prophages are classified within the proposed parabacteroides prophage subfamily. In addition, parabacteroides prophages related to the Microviridae have been identified (Ref). Here we identified 7 CGM genomes similar to the parabacteroides prophage representative sequences, suggesting the possibility that ssDNA prophages exist in this system. Although these sequences are in the minority in regards to the overall CGM diversity, it suggests that lysogeny may be possible if not common in the *Ciona* gut. It is possible that some *Microviridae* phages become temperate (integrating into a bacterial host genome) in the *Ciona* gut and that gut clearing may result in prophage induction, though further studies are necessary.

There remains a dearth of knowledge concerning the distribution and structure of viral communities among microbiomes over time. Many studies focus on the structure of the gut virome in response to disease (Kramná *et al.* 2015, Reyes *et al.* 2015, Guo *et al.* 2017, Zhao *et al.* 2017) though fewer studies have assessed normal changes of a healthy virome (Reyes *et al.* 2010, Lim *et al.* 2015). The *Ciona* gut also has more *Microviridae* diversity than any environmental (Tucker *et al.* 2011, Labonté and Suttle 2013, Labonté and Suttle 2013, Hopkins *et al.* 2014, Bryson *et al.* 2015, Labonté *et al.* 2015, Quaiser *et al.* 2015, Zhong *et al.* 2015, Yu *et al.* 2018) or gut study (Reyes *et al.* 2010, Krupovic *et al.* 2011, Roux *et al.* 2012, Reyes *et al.* 2013, Waller *et al.* 2014, Lim *et al.* 2015, Reyes *et al.* 2015, Tikhe and Husseneder 2017, D'arc *et al.* 2018, McCann *et al.* 2018) to date. Many of the studies examining animal guts have defined the structure of the viral community but have not probed the diversity of ssDNA phages

present. A few recent studies have looked at *Microviridae* diversity in human feces. For example, one in depth study looking at patients with coronary heart disease found 12 *Gokushovirinae* genomes and 2 *Microviridae* genomes that did not group with any known subfamily (Guo *et al.* 2017). Other studies using human feces have identified the presence of *Gokushovirinae*; however, these were all done as part of larger studies and not explored deeply (Waller *et al.* 2014, Reyes *et al.* 2015). *Microviridae* diversity has recently been assessed in the gut of termites. This study found 12 *Microviridae* genomes, 2 of which were *Gokushovirinae*, 3 did not group with any reference genomes and 7 were proposed to belong to a new subfamily *Sukshmavirinae* (Tikhe and Husseneder 2017).

The 258 unique *Microviridae* genomes described in this study of *Ciona* guts originate from only 5 clear animals and 5 full animals, which is a remarkable diversity for any single study. These findings suggest that at least some of these phages, some of which may be infecting intracellular bacteria, could have native hosts colonizing or infecting the *Ciona* gut. As a filter-feeding organism that concentrates organic material from seawater, *Ciona* provides unique opportunities to explore questions about *Microviridae* within the gut environment. This is particularly true as the microbiome (Dishaw *et al.* 2014) can be manipulated and tightly controlled by rearing the animals germ-free (Leigh *et al.* 2016). Juvenile *Ciona* are small enough that dozens to hundreds of transparent juveniles can be reared on small tissue culture dishes facilitating experimental manipulations. Some straightforward hypotheses to explore are that some of these viruses may be vertically transferred or could be induced through starvation. A filter-feeding system like *Ciona* affords unique opportunities to address hypothesis-driven questions and possibly develop the first model for understanding the host distribution and biology of the *Microviridae* in animal guts.

References

Angly, F. E., B. Felts, M. Breitbart, P. Salamon, R. A. Edwards, C. Carlson, A. M. Chan, M. Haynes, S. Kelley and H. Liu (2006). "The marine viromes of four oceanic regions." PLoS Biology **4**(11): e368.

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham and A. D. Prjibelski (2012). "Spades: A new genome assembly algorithm and its applications to single-cell sequencing." Journal of computational biology **19**(5): 455-477.

Basso, M. F., J. C. F. da Silva, T. V. M. Fajardo, E. P. B. Fontes and F. M. Zerbini (2015). "A novel, highly divergent ssdna virus identified in brazil infecting apple, pear and grapevine." Virus Research **210**: 27-33.

Bench, S. R., T. E. Hanson, K. E. Williamson, D. Ghosh, M. Radosovich, K. Wang and K. E. Wommack (2007). "Metagenomic characterization of chesapeake bay virioplankton." Applied and Environmental Microbiology **73**(23): 7629-7641.

Bolduc, B., K. Youens-Clark, S. Roux, B. L. Hurwitz and M. B. Sullivan (2017). "Ivirus: Facilitating new insights in viral ecology with software and community data sets imbedded in a cyberinfrastructure." The ISME Journal **11**(1): 7.

Bolger, A. M., M. Lohse and B. Usadel (2014). "Trimmomatic: A flexible trimmer for illumina sequence data." Bioinformatics **30**(15): 2114-2120.

Bordenstein, S. R. and K. R. Theis (2015). "Host biology in light of the microbiome: Ten principles of holobionts and hologenomes." PLoS Biol **13**(8): e1002226.

Breitbart, M. (2011). "Marine viruses: Truth or dare."

Breitbart, M., M. Haynes, S. Kelley, F. Angly, R. A. Edwards, B. Felts, J. M. Mahaffy, J. Mueller, J. Nulton and S. Rayhawk (2008). "Viral diversity and dynamics in an infant gut." Research in Microbiology **159**(5): 367-373.

Breitbart, M., I. Hewson, B. Felts, J. M. Mahaffy, J. Nulton, P. Salamon and F. Rohwer (2003). "Metagenomic analyses of an uncultured viral community from human feces." Journal of Bacteriology **185**(20): 6220-6223.

Breitbart, M., P. Salamon, B. Andresen, J. M. Mahaffy, A. M. Segall, D. Mead, F. Azam and F. Rohwer (2002). "Genomic analysis of uncultured marine viral communities." Proceedings of the National Academy of Sciences **99**(22): 14250-14255.

Bryson, S. J., A. R. Thurber, A. Correa, V. J. Orphan and R. Vega Thurber (2015). "A novel sister clade to the enterobacteria microviruses (family *Microviridae*) identified in methane seep sediments." Environmental microbiology **17**(10): 3708-3721.

Carding, S. R., N. Davis and L. Hoyles (2017). "The human intestinal virome in health and disease." Alimentary Pharmacology & Therapeutics **46**(9): 800-815.

Chen, H. and P. C. Boutros (2011). "Venndiagram: A package for the generation of highly-customizable venn and euler diagrams in r." BMC bioinformatics **12**(1): 35.

Cherwa, J. E. and B. A. Fane (2001). *Microviridae*: Microviruses and gokushoviruses. Els, John Wiley & Sons, Ltd.

D'arc, M., C. Furtado, J. D. Siqueira, H. N. Seuánez, A. Ayouba, M. Peeters and M. A. Soares (2018). "Assessment of the gorilla gut virome in association with natural simian immunodeficiency virus infection." Retrovirology **15**(1): 19.

Dahiya, D. K. (2017). The gut virome: A neglected actor in colon cancer pathogenesis, Future Medicine.

Dehal, P., Y. Satou, R. K. Campbell, J. Chapman, B. Degnan, A. De Tomaso, B. Davidson, A. Di Gregorio, M. Gelpke and D. M. Goodstein (2002). "The draft genome of *Ciona intestinalis*: Insights into chordate and vertebrate origins." Science **298**(5601): 2157-2167.

Desnues, C., B. Rodriguez-Brito, S. Rayhawk, S. Kelley, T. Tran, M. Haynes, H. Liu, M. Furlan, L. Wegley and B. Chau (2008). "Biodiversity and biogeography of phages in modern stromatolites and thrombolites." Nature **452**(7185): 340.

Dishaw, L. J., J. Flores-Torres, S. Lax, K. Gemayel, B. Leigh, D. Melillo, M. G. Mueller, L. Natale, I. Zucchetti, R. De Santis, M. R. Pinto, G. W. Litman and J. A. Gilbert (2014). "The gut of geographically disparate *Ciona intestinalis* harbors a core microbiota." Plos One **9**(4): 8.

Doore, S. M. and B. A. Fane (2016). "The *Microviridae*: Diversity, assembly, and experimental evolution." Virology **491**: 45-55.

Edgar, R. C. (2004). "Muscle: Multiple sequence alignment with high accuracy and high throughput." Nucleic Acids Research **32**(5): 1792-1797.

Goff, S. A., M. Vaughn, S. McKay, E. Lyons, A. E. Stapleton, D. Gessler, N. Matasci, L. Wang, M. Hanlon and A. Lenards (2011). "The iplant collaborative: Cyberinfrastructure for plant biology." Frontiers in Plant Science **2**: 34.

Grzenia, D. L., J. O. Carlson and S. R. Wickramasinghe (2008). "Tangential flow filtration for virus purification." Journal of Membrane Science **321**(2): 373-380.

Guo, L., X. Hua, W. Zhang, S. Yang, Q. Shen, H. Hu, J. Li, Z. Liu, X. Wang and H. Wang (2017). "Viral metagenomics analysis of feces from coronary heart disease patients reveals the genetic diversity of the *Microviridae*." Virologica Sinica **2**(32): 130-138.

Han, L.-L., D.-T. Yu, L.-M. Zhang, J.-P. Shen and J.-Z. He (2017). "Genetic and functional diversity of ubiquitous DNA viruses in selected chinese agricultural soils." Scientific Reports **7**: 45142.

Hopkins, M., S. Kailasan, A. Cohen, S. Roux, K. P. Tucker, A. Shevenell, M. Agbandje-McKenna and M. Breitbart (2014). "Diversity of environmental single-stranded DNA phages revealed by pcr amplification of the partial major capsid protein." The ISME Journal **8**(10): 2093-2103.

Huson, D. H., S. Beier, I. Flade, A. Górska, M. El-Hadidi, S. Mitra, H.-J. Ruscheweyh and R. Tappu (2016). "Megan community edition-interactive exploration and analysis of large-scale microbiome sequencing data." PLoS Computational Biology **12**(6): e1004957.

Jørgensen, T. S., Z. Xu, M. A. Hansen, S. J. Sørensen and L. H. Hansen (2014). "Hundreds of circular novel plasmids and DNA elements identified in a rat cecum metamobilome." PLoS One **9**(2): e87924.

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz and C. Duran (2012). "Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data." Bioinformatics **28**(12): 1647-1649.

Kim, K.-H. and J.-W. Bae (2011). "Amplification methods bias metagenomic libraries of uncultured single-stranded and double-stranded DNA viruses." Applied and Environmental Microbiology **77**(21): 7663-7668.

Kim, K.-H., H.-W. Chang, Y.-D. Nam, S. W. Roh, M.-S. Kim, Y. Sung, C. O. Jeon, H.-M. Oh and J.-W. Bae (2008). "Amplification of uncultured single-stranded DNA viruses from rice paddy soil." Applied and Environmental Microbiology **74**(19): 5975-5985.

Kim, M.-S., E.-J. Park, S. W. Roh and J.-W. Bae (2011). "Diversity and abundance of single-stranded DNA viruses in human feces." Applied and Environmental Microbiology **77**(22): 8062-8070.

Koskella, B. and M. A. Brockhurst (2014). "Bacteria–phage coevolution as a driver of ecological and evolutionary processes in microbial communities." FEMS Microbiology Reviews **38**(5): 916-931.

Kramná, L., K. Kolářová, S. Oikarinen, J.-P. Pursiheimo, J. Ilonen, O. Simell, M. Knip, R. Veijola, H. Hyöty and O. Cinek (2015). "Gut virome sequencing in children with early islet autoimmunity." Diabetes Care **38**(5): 930-933.

Krupovic, M. and P. Forterre (2011). "*Microviridae* goes temperate: Microvirus-related proviruses reside in the genomes of *Bacteroidetes*." PLoS One **6**(5): e19893.

Krupovic, M., D. Prangishvili, R. W. Hendrix and D. H. Bamford (2011). "Genomics of bacterial and archaeal viruses: Dynamics within the prokaryotic virosphere." Microbiology and Molecular Biology Reviews **75**(4): 610-635.

Labonté, J. and C. Suttle (2013). "Metagenomic and whole-genome analysis reveals new lineages of gokushoviruses and biogeographic separation in the sea." Frontiers in Microbiology **4**(404).

Labonté, J. M., S. J. Hallam and C. A. Suttle (2015). "Previously unknown evolutionary groups dominate the ssdna gokushoviruses in oxic and anoxic waters of a coastal marine environment." Frontiers in microbiology **6**: 315.

Labonté, J. M. and C. A. Suttle (2013). "Previously unknown and highly divergent ssdna viruses populate the oceans." The ISME Journal **7**(11): 2169.

Lefkowitz, E. J., D. M. Dempsey, R. C. Hendrickson, R. J. Orton, S. G. Siddell and D. B. Smith (2017). "Virus taxonomy: The database of the international committee on taxonomy of viruses (ictv)." Nucleic Acids Research: gkx932.

Lefort, V., J.-E. Longueville and O. Gascuel (2017). "Sms: Smart model selection in phyml." Molecular Biology and Evolution **34**(9): 2422-2424.

Leigh, B. A., A. Djurhuus, M. Breitbart and L. J. Dishaw (2018). "The gut virome of the protochordate model organism, *Ciona intestinalis* subtype a." Virus Research **244**: 137-146.

Leigh, B. A., A. Liberti and L. J. Dishaw (2016). "Generation of germ-free *Ciona intestinalis* for studies of gut-microbe interactions." Frontiers in Microbiology **7**.

Lim, E. S., Y. Zhou, G. Zhao, I. K. Bauer, L. Droit, I. M. Ndao, B. B. Warner, P. I. Tarr, D. Wang and L. R. Holtz (2015). "Early life dynamics of the human gut virome and bacterial microbiome in infants." Nature Medicine **21**(10): 1228.

López-Bueno, A., J. Tamames, D. Velázquez, A. Moya, A. Quesada and A. Alcamí (2009). "High diversity of the viral community from an Antarctic lake." Science **326**(5954): 858-861.

Manrique, P., B. Bolduc, S. T. Walk, J. van der Oost, W. M. de Vos and M. J. Young (2016). "Healthy human gut phageome." Proceedings of the National Academy of Sciences **113**(37): 10400-10405.

Manrique, P., M. Dills and M. J. Young (2017). "The human gut phage community and its implications for health and disease." Viruses **9**(6): 141.

McCann, A., F. J. Ryan, S. R. Stockdale, M. Dalmasso, T. Blake, C. A. Ryan, C. Stanton, S. Mills, P. R. Ross and C. Hill (2018). "Viromes of one year old infants reveal the impact of birth mode on microbiome diversity." PeerJ **6**: e4694.

Mills, S., F. Shanahan, C. Stanton, C. Hill, A. Coffey and R. P. Ross (2013). "Movers and shakers: Influence of bacteriophages in shaping the mammalian gut microbiota." Gut Microbes **4**(1): 4-16.

Minot, S., A. Bryson, C. Chehoud, G. D. Wu, J. D. Lewis and F. D. Bushman (2013). "Rapid evolution of the human gut virome." Proceedings of the National Academy of Sciences **110**(30): 12450-12455.

Minot, S., R. Sinha, J. Chen, H. Li, S. A. Keilbaugh, G. D. Wu, J. D. Lewis and F. D. Bushman (2011). "The human gut virome: Inter-individual variation and dynamic response to diet." Genome research **21**(10): 1616-1625.

Mirzaei, M. K. and C. F. Maurice (2017). "Ménage à trois in the human gut: Interactions between host, bacteria and phages." Nature Reviews Microbiology **15**(7): 397.

Moreno, P. S., J. Wagner, C. S. Mansfield, M. Stevens, J. R. Gilkerson and C. D. Kirkwood (2017). "Characterisation of the canine faecal virome in healthy dogs and dogs with acute diarrhoea using shotgun metagenomics." PloS One **12**(6): e0178433.

Muhire, B. M., A. Varsani and D. P. Martin (2014). "Sdt: A virus classification tool based on pairwise sequence alignment and identity calculation." PloS one **9**(9): e108277.

Nicholson, J. K., E. Holmes, J. Kinross, R. Burcelin, G. Gibson, W. Jia and S. Pettersson (2012). "Host-gut microbiota metabolic interactions." Science: 1223813.

Norman, J. M., S. A. Handley, M. T. Baldridge, L. Droit, C. Y. Liu, B. C. Keller, A. Kambal, C. L. Monaco, G. Zhao and P. Fleshner (2015). "Disease-specific alterations in the enteric virome in inflammatory bowel disease." Cell **160**(3): 447-460.

O'Toole, P. W. and I. B. Jeffery (2015). "Gut microbiota and aging." Science **350**(6265): 1214-1215.

Ogilvie, L. A. and B. V. Jones (2015). "The human gut virome: A multifaceted majority." Frontiers in microbiology **6**: 918.

Oksanen, J., R. Kindt, P. Legendre, B. O'Hara, M. H. H. Stevens, M. J. Oksanen and M. Suggests (2007). "The vegan package." Community ecology package **10**: 631-637.

Pearson, V. M., S. B. Caudle and D. R. Rokyta (2016). "Viral recombination blurs taxonomic lines: Examination of single-stranded DNA viruses in a wastewater treatment plant." PeerJ **4**: e2585.

Quaiser, A., A. Dufresne, F. Ballaud, S. Roux, Y. Zivanovic, J. Colombet, T. Sime-Ngando and A.-J. Francez (2015). "Diversity and comparative genomics of *Microviridae* in sphagnum-dominated peatlands." Frontiers in Microbiology **6**(375).

Reavy, B., M. M. Swanson, P. J. Cock, L. Dawson, T. E. Freitag, B. K. Singh, L. Torrance, A. R. Mushegian and M. Taliansky (2015). "Distinct circular single-stranded DNA viruses exist in different soil types." Applied and Environmental Microbiology **81**(12): 3934-3945.

Reyes, A., L. V. Blanton, S. Cao, G. Zhao, M. Manary, I. Trehan, M. I. Smith, D. Wang, H. W. Virgin and F. Rohwer (2015). "Gut DNA viromes of malawian twins discordant for severe acute malnutrition." Proceedings of the National Academy of Sciences **112**(38): 11941-11946.

Reyes, A., M. Haynes, N. Hanson, F. E. Angly, A. C. Heath, F. Rohwer and J. I. Gordon (2010). "Viruses in the faecal microbiota of monozygotic twins and their mothers." Nature **466**(7304): 334-338.

Reyes, A., N. P. Semenkovich, K. Whiteson, F. Rohwer and J. I. Gordon (2012). "Going viral: Next-generation sequencing applied to phage populations in the human gut." Nature Reviews Microbiology **10**(9): 607-617.

Reyes, A., M. Wu, N. P. McNulty, F. L. Rohwer and J. I. Gordon (2013). "Gnotobiotic mouse model of phage–bacterial host dynamics in the human gut." Proceedings of the National Academy of Sciences **110**(50): 20236-20241.

Rohwer, F. and R. V. Thurber (2009). "Viruses manipulate the marine environment." Nature **459**(7244): 207.

Rosario, K., A. Dayaram, M. Marinov, J. Ware, S. Kraberger, D. Stainton, M. Breitbart and A. Varsani (2012). "Diverse circular ssdna viruses discovered in dragonflies (odonata: Epiprocta)." Journal of General Virology **93**(12): 2668-2681.

Rosario, K., C. Nilsson, Y. W. Lim, Y. Ruan and M. Breitbart (2009). "Metagenomic analysis of viruses in reclaimed water." Environmental microbiology **11**(11): 2806-2820.

Roux, S., F. Enault, A. Robin, V. Ravet, S. Personnic, S. Theil, J. Colombet, T. Sime-Ngando and D. Debroas (2012). "Assessing the diversity and specificity of two freshwater viral communities through metagenomics." PloS one **7**(3): e33641.

Roux, S., M. Faubladier, A. Mahul, N. Paulhe, A. Bernard, D. Debroas and F. Enault (2011). "Metavir: A web server dedicated to virome analysis." Bioinformatics **27**(21): 3074-3075.

Roux, S., M. Krupovic, A. Poulet, D. Debroas and F. Enault (2012). "Evolution and diversity of the *Microviridae* viral family through a collection of 81 new complete genomes assembled from virome reads." PloS One **7**(7): e40418.

Roux, S., N. E. Solonenko, V. T. Dang, B. T. Poulos, S. M. Schwenck, D. B. Goldsmith, M. L. Coleman, M. Breitbart and M. B. Sullivan (2016). "Towards quantitative viromics for both double-stranded and single-stranded DNA viruses." PeerJ **4**: e2777.

Roux, S., J. Tournayre, A. Mahul, D. Debroas and F. Enault (2014). "Metavir 2: New tools for viral metagenome comparison and assembled virome analysis." BMC bioinformatics **15**(1): 76.

Sakamoto, M. and Y. Benno (2006). "Reclassification of *Bacteroides distasonis*, *Bacteroides goldsteinii* and *Bacteroides merdae* as *Parabacteroides distasonis* gen. Nov., comb. Nov., parabacteroides goldsteinii comb. Nov. And parabacteroides merdae comb. Nov." International Journal of Systematic and Evolutionary Microbiology **56**(7): 1599-1605.

Sanders, J. G., A. C. Beichman, J. Roman, J. J. Scott, D. Emerson, J. J. McCarthy and P. R. Girguis (2015). "Baleen whales host a unique gut microbiome with similarities to both carnivores and herbivores." Nature Communications **6**: 8285.

Santiago-Rodriguez, T. M., G. Fornaciari, S. Luciani, S. E. Dowd, G. A. Toranzos, I. Marota and R. J. Cano (2015). "Natural mummification of the human gut preserves bacteriophage DNA." FEMS microbiology letters **363**(1): fnv219.

Satoh, N., Y. Satou, B. Davidson and M. Levine (2003). "*Ciona intestinalis*: An emerging model for whole-genome analyses." TRENDS in Genetics **19**(7): 376-381.

Scarpellini, E., G. Ianiro, F. Attili, C. Bassanelli, A. De Santis and A. Gasbarrini (2015). "The human gut microbiota and virome: Potential therapeutic implications." Digestive and Liver Disease **47**(12): 1007-1012.

Shapiro, O. H., A. Kushmaro and A. Brenner (2010). "Bacteriophage predation regulates microbial abundance and diversity in a full-scale bioreactor treating industrial wastewater." The ISME journal **4**(3): 327.

Smith, R. J., T. C. Jeffries, B. Roudnew, J. R. Seymour, A. J. Fitch, K. L. Simons, P. G. Speck, K. Newton, M. H. Brown and J. G. Mitchell (2013). "Confined aquifers as viral reservoirs." Environmental microbiology reports **5**(5): 725-730.

Soverini, M., S. Quercia, B. Biancani, S. Furlati, S. Turroni, E. Biagi, C. Consolandi, C. Peano, M. Severgnini and S. Rampelli (2016). "The bottlenose dolphin (tursiops truncatus) faecal microbiota." FEMS microbiology ecology **92**(4): fiw055.

Stöver, B. C. and K. F. Müller (2010). "Treegraph 2: Combining and visualizing evidence from different phylogenetic analyses." BMC bioinformatics **11**(1): 7.

Suttle, C. A. (2007). "Marine viruses—major players in the global ecosystem." Nature Reviews Microbiology **5**(10): 801-812.

Team, R. C. (2014). R: A language and environment for statistical computing. Vienna, Austria: R foundation for statistical computing; 2014.

Theis, K. R., N. M. Dheilly, J. L. Klassen, R. M. Brucker, J. F. Baines, T. C. Bosch, J. F. Cryan, S. F. Gilbert, C. J. Goodnight and E. A. Lloyd (2016). "Getting the hologenome concept right: An eco-evolutionary framework for hosts and their microbiomes." Msystems **1**(2): e00028-00016.

Thurber, R. V., M. Haynes, M. Breitbart, L. Wegley and F. Rohwer (2009). "Laboratory procedures to generate viral metagenomes." Nature Protocols **4**(4): 470-483.

Thurber, R. V., J. P. Payet, A. R. Thurber and A. M. Correa (2017). "Virus–host interactions and their roles in coral reef health and disease." Nature Reviews Microbiology **15**(4): 205.

Tikhe, C. V. and C. Husseneder (2017). "Metavirome sequencing of the termite gut reveals the presence of a diverse bacteriophage community." Frontiers in Microbiology **8**: 2548.

Tucker, K. P., R. Parsons, E. M. Symonds and M. Breitbart (2011). "Diversity and distribution of single-stranded DNA phages in the North Atlantic ocean." The ISME Journal **5**(5): 822-830.

Turnbaugh, P. J., R. E. Ley, M. Hamady, C. M. Fraser-Liggett, R. Knight and J. I. Gordon (2007). "The human microbiome project." Nature **449**(7164): 804.

Waller, A. S., T. Yamada, D. M. Kristensen, J. R. Kultima, S. Sunagawa, E. V. Koonin and P. Bork (2014). "Classification and quantification of bacteriophage taxa in human gut metagenomes." The ISME Journal **8**(7): 1391.

Walters, M., M. Bawuro, A. Christopher, A. Knight, S. Kraberger, D. Stainton, H. Chapman and A. Varsani (2017). "Novel single-stranded DNA virus genomes recovered from chimpanzee feces sampled from the mambilla plateau in Nigeria." Genome Announcements **5**(9): e01715-01716.

Weisburg, W. G., S. M. Barns, D. A. Pelletier and D. J. Lane (1991). "16s ribosomal DNA amplification for phylogenetic study." Journal of Bacteriology **173**(2): 697-703.

Yatsunenko, T., F. E. Rey, M. J. Manary, I. Trehan, M. G. Dominguez-Bello, M. Contreras, M. Magris, G. Hidalgo, R. N. Baldassano and A. P. Anokhin (2012). "Human gut microbiome viewed across age and geography." Nature **486**(7402): 222.

Yoshida, M., T. Mochizuki, S.-I. Urayama, Y. Yoshida-Takashima, S. Nishi, M. Hirai, H. Nomaki, Y. Takaki, T. Nunoura and K. Takai (2018). "Quantitative viral community DNA analysis reveals the dominance of single-stranded DNA viruses in offshore upper bathyal sediment from Tohoku, Japan." Frontiers in Microbiology **9**: 75.

Yoshida, M., Y. Takaki, M. Eitoku, T. Nunoura and K. Takai (2013). "Metagenomic analysis of viral communities in (hado) pelagic sediments." PloS one **8**(2): e57271.

Yu, D.-T., L.-L. Han, L.-M. Zhang and J.-Z. He (2018). "Diversity and distribution characteristics of viruses in soils of a marine-terrestrial ecotone in east China." Microbial Ecology **75**(2): 375-386.

Yutin, N., K. S. Makarova, A. B. Gussow, M. Krupovic, A. Segall, R. A. Edwards and E. V. Koonin (2018). "Discovery of an expansive bacteriophage family that includes the most abundant viruses from the human gut." Nature Microbiology **3**(1): 38.

Zhang, T., M. Breitbart, W. H. Lee, J.-Q. Run, C. L. Wei, S. W. L. Soh, M. L. Hibberd, E. T. Liu, F. Rohwer and Y. Ruan (2005). "Rna viral community in human feces: Prevalence of plant pathogenic viruses." PLoS Biology **4**(1): e3.

Zhao, G., T. Vatanen, L. Droit, A. Park, A. D. Kostic, T. W. Poon, H. Vlamakis, H. Siljander, T. Härkönen and A.-M. Hämäläinen (2017). "Intestinal virome changes precede autoimmunity in type i diabetes-susceptible children." Proceedings of the National Academy of Sciences: 201706359.

Zhong, X., B. Guidoni, L. Jacas and S. Jacquet (2015). "Structure and diversity of ssdna *Microviridae* viruses in two peri-alpine lakes (Annecy and Bourget, France)." Research in microbiology **166**(8): 644-654.