2018

Unprecedented Diversity of ssDNA Phages from the Family Microviridae Detected within the Gut of a Protochordate Model Organism (Ciona robusta)

Alexandria Creasy  
*University of South Florida*

Karyna Rosario  
*University of South Florida*

Brittany Leigh  
*University of South Florida*

Larry Dishaw  
*University of South Florida*

Mya Breitbart  
*University of South Florida, mya@usf.edu*

Follow this and additional works at: [https://scholarcommons.usf.edu/msc_facpub](https://scholarcommons.usf.edu/msc_facpub)

🔗 Part of the *Life Sciences Commons*

---

**Scholar Commons Citation**

Creasy, Alexandria; Rosario, Karyna; Leigh, Brittany; Dishaw, Larry; and Breitbart, Mya, "Unprecedented Diversity of ssDNA Phages from the Family Microviridae Detected within the Gut of a Protochordate Model Organism (Ciona robusta)" (2018). *Marine Science Faculty Publications*. 671.  
[https://scholarcommons.usf.edu/msc_facpub/671](https://scholarcommons.usf.edu/msc_facpub/671)

This Article is brought to you for free and open access by the College of Marine Science at Scholar Commons. It has been accepted for inclusion in Marine Science Faculty Publications by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.
Unprecedented Diversity of ssDNA Phages from the Family Microviridae Detected within the Gut of a Protochordate Model Organism (Ciona robusta)

Alexandria Creasy, Karyna Rosario, Brittany A. Leigh, Larry J. Dishaw, and Mya Breitbart

1 College of Marine Science, University of South Florida, St. Petersburg, FL 33701, USA; creasy@mail.usf.edu (A.C.); krosari2@mail.usf.edu (K.R.), bleigh@mail.usf.edu or brittany.a.leigh@vanderbilt.edu (B.A.L.)
2 Department of Pediatrics, Children’s Research Institute, University of South Florida, St. Petersburg, FL 33701, USA
* Correspondence: ldishaw@health.usf.edu (L.J.D.); mya@usf.edu (M.B.); Tel.: +1-727-553-3601 (L.J.D.); +1-727-553-3520 (M.B.)
† Current address: Department of Biological Sciences, Vanderbilt University, Nashville, TN 37235, USA.

Received: 7 July 2018; Accepted: 27 July 2018; Published: 31 July 2018

Abstract: Phages (viruses that infect bacteria) play important roles in the gut ecosystem through infection of bacterial hosts, yet the gut virome remains poorly characterized. Mammalian gut viromes are dominated by double-stranded DNA (dsDNA) phages belonging to the order Caudovirales and single-stranded DNA (ssDNA) phages belonging to the family Microviridae. Since the relative proportion of each of these phage groups appears to correlate with age and health status in humans, it is critical to understand both ssDNA and dsDNA phages in the gut. Building upon prior research describing dsDNA viruses in the gut of Ciona robusta, a marine invertebrate model system used to study gut microbial interactions, this study investigated ssDNA phages found in the Ciona gut. We identified 258 Microviridae genomes, which were dominated by novel members of the Gokushovirinae subfamily, but also represented several proposed phylogenetic groups (Alpavirinae, Aravirinae, Group D, Parabacteroides prophages, and Pequeñovirus) and a novel group. Comparative analyses between Ciona specimens with full and cleared guts, as well as the surrounding water, indicated that Ciona retains a distinct and highly diverse community of ssDNA phages. This study significantly expands the known diversity within the Microviridae family and demonstrates the promise of Ciona as a model system for investigating their role in animal health.

Keywords: phage; ssDNA; Microviridae; virome; microbiome; gut; Ciona; invertebrate

1. 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 [1]. The complex, dynamic relationship between an animal host and its microbial associates impacts many aspects of host physiology, including the procurement of nutrients and metabolic output [2,3]. In the nutrient- and mucus-rich gut of animals, the bacterial component of the microbiome can contribute orders of magnitude more gene products to host physiology than the host genome itself [4]. Although significant attention has been given to the metabolic contributions from the cellular component of the microbiome (i.e., bacteria, archaea, and fungi), it is now recognized that viruses also play important roles in the holobiont, both directly through infecting animal host cells, and indirectly through mediating microbiome dynamics.
Viruses present in gut communities are dominated by phages, which are viruses that infect bacteria [5–11]. The abundances of phages and bacteria are approximately equal in the human intestine [6,7,12], which contrasts with ratios observed in environmental samples where phages are typically ten times as abundant as bacteria [13]. Since phages can have dramatic influences on the structure and function of bacterial communities through host-specific lysis and horizontal gene transfer [13–15], phage-bacterial interactions appear crucial in all environments studied to date [15–19], including communities that are found in close association with animals [4,20–24]. 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 impacts within the gut environment [6,8–10,25–30].

A variety of factors may influence the abundance, diversity, and ecological roles of phages in a complex ecosystem like the gut [31,32]. However, characterization of phage community composition is difficult due to a lack of universal gene markers within phage genomes [16]. Viral metagenomics, where the collective viral nucleic acids from a given sample are sequenced, is an efficient alternative approach for exploring the gut virome [5,12]. Although both single-stranded DNA (ssDNA) and double-stranded DNA (dsDNA) phages have been identified in animal guts [27,33], the diversity of dsDNA phages remains the best characterized due to their large representation in culture collections and genomic sequence databases. In addition, standardly used library construction kits (e.g., the NexteraXT DNA library preparation kit, Illumina, San Diego, CA, USA) and viromic studies using linker-amplified shotgun sequencing approaches bias against ssDNA viruses [34]. 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 [6,7,12,33,35–47].

Studies investigating microbial communities in the human gut have shown that the microbiome structure seen in healthy newborns develops and changes dramatically within the first 2–3 years of life from a nearly sterile gut environment to a dynamic community maintained throughout adulthood [27]. Human gut phage communities are dominated by dsDNA phages from the order Caudovirales and members of the ssDNA phage family Microviridae [27,48,49]. Infant viromes appear to contain the highest phage richness at birth, which decreases with age [27,50]. While Caudovirales dominate in the earliest samples, the composition of the virome changes over time, with an increased relative abundance and richness of Microviridae by 2 years of age [27]. Gut phage communities often change in composition during disease states, such as inflammatory bowel disease, where the Microviridae:Caudovirales ratio decreases [51]. Despite the fact that several recent studies have highlighted phages belonging to the Microviridae as important gut virome components associated with an individual’s health state [12,27,33,37,51,52], little is known about the roles and dynamics of Microviridae in this complex ecosystem.

Studies leveraging simpler model systems to study the gut virome can enable 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 [53] with a sequenced genome [54]. Because Ciona is a filter feeding organism, its gut is a microcosm of microbial interactions that experience vast and continuous exposure to the large microbial and viral diversity found in seawater. Previous efforts have identified remarkable stability among some elements of the microbiome, which are distinct from surrounding seawater [20,55]. Ciona has structurally distinct gut compartments (stomach, midgut, hindgut) and maintains core bacterial species, some of which exhibit compartmentalization [20]. A recent effort characterizing the dsDNA Ciona gut virome revealed the dominance of tailed phages belonging to the order Caudovirales, as well as a diversity of eukaryotic dsDNA viruses. There was also evidence for compartmentalization of some components of the dsDNA virome, although to a lesser extent than seen for the corresponding bacterial communities [55]. 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 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 the identified ssDNA phages were unique to Ciona or if they were also found in water samples, and (3) determine if different gut compartments contained distinct ssDNA phage assemblages. We show that Ciona harbors a highly diverse community of ssDNA phages that is distinct from the water column and demonstrates some compartmentalization within the gut, but with the majority of genomes found in multiple gut compartments. The novel viral genomes reported here significantly broaden the known diversity of the Microviridae family, with a particular expansion in the subfamily Gokushovirinae and detection of a novel phylogenetic group.

2. Materials and Methods

2.1. 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 viral diversity. Animal guts were sampled from Ciona robusta harvested near San Diego, CA (M-Rep, Carlsbad, CA, USA). Ciona robusta is an invertebrate that is not regulated or protected by environmental agencies in the USA. M-Rep maintains a valid collection license through the state of California that allows them to ship to research institutions throughout the USA. Research at the University of South Florida was approved by biosafety protocols 1199 IA and 1351E. Upon sample arrival in Florida, ten animals were selected at random; five specimens were placed into virus-free 100 kD-filtered artificial seawater [56] to clear guts of dietary contents (water changed every 4 h for 24 h). 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 (6000 × g for 10 min), the supernatant was filtered through a 0.22 µm pore size Sterivex filter (Merck Millipore, Burlington, MA, USA), 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) where the animals were collected, 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 pore size 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 [56] and collecting the 1.2–1.5 g/mL fraction with a sterile syringe and needle into a 2 mL sterile tube. To remove potential bacteria or extracellular vesicles still present in the sample, chloroform (final concentration 20% v/v) was added to the viral fraction and incubated at room temperature for 10 min. Samples were then centrifuged for 30 s at maximum speed (20,000 × g) 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 20 mM final concentration of EDTA pH 8.0. Purified VLP samples were tested to rule out bacterial contamination by PCR amplification of the 16S rRNA gene using primers 27F and 1492R [57] 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 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 determine the DNA concentration and amplification was verified with 1%
agarose gel electrophoresis. Final, amplified products were cleaned via the MinElute PCR Purification Kit (Qiagen, Inc.). DNA quality and quantity were assessed using the BioAnalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA). Sequencing was performed on an Illumina MiSeq platform generating mate-pair (2 × 250 bp) libraries (Operon, Eurofins MWG Operon LLC, Huntsville, AL, USA). Sequences were analyzed in the CyVerse Cyberinfrastructure using different bioinformatics applications (Apps) [58]. Briefly, raw sequences were trimmed based on quality scores using the Trimmomatic App version 0.35.0 [59] and quality-filtered sequences were then assembled using the SPAdes App version 3.6.0 [60]. Contigs were screened with the VirSorter App to detect potential viral sequences. VirSorter outputs were uploaded to MetaVir [61] (IDs 7811 (SF), 8143 (SC), 7815 (MF), 7814 (MC), 7812 (HF), 7910 (HC), 7816 (CB), 7819 (MB)).

2.2. Identification and Annotation of Microviridae Genomes

Since members of the Microviridae family have circular genomes, circular contig sequences were identified using MetaVir [62]. 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 (Jonathan Vincent and Francois Enault, Université Clermont Auvergne). Contig sequences 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 [63] to identify sequences related to the Microviridae. Microviridae-related sequences were then manually trimmed to unit length genomes by identifying repeated sequences, and unit-length genomes were annotated using Geneious v10.1.3 [64]. 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). Whenever possible, ORFs were annotated based on phage proteins (PHA source database) found in the NCBI conserved domain database with nomenclature following Cherwa and Fane [65]. All genomes were manually edited to begin at the start codon of the MCP. Genome-wide and MCP pairwise identities were calculated using the sequence demarcation tool (SDT) v1.2 [66]. Genomes were clustered at 95% nucleic acid identity based on current species demarcation cutoffs for phages [67]. Complete genome sequences are available under Genbank Accession Numbers MH572269–MH572526.

2.3. Genome Comparisons and Phylogenetic Analysis

Reference MCP amino acid sequences, including VP1 (subfamily Gokushovirinae) and Protein F (subfamily Bullavirinae), were collected from GenBank. These reference sequences also contained select sequences related to Microviridae that were identified from metagenomes [37,68–76] and those integrated into bacterial genomes [36] (see Supplementary Table S1 for details). Reference MCP sequences were aligned with sequences identified in Ciona guts using MUSCLE [77] as implemented in Geneious v10.1.3 [64]. A maximum likelihood tree was then created using PhyML with aLRT-like probabilities for branch support [78] 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 [79].

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 [80]. For this purpose, Bowtiebatch v1.0.0 and Read2Ref v1.0.1 were used with default parameters to map reads from each gut compartment against the 258 Microviridae genomes, requiring reads to cover >75% of the genome to consider a genome present in a compartment [80]. A binary matrix (presence/absence) was used to assess the number of shared Microviridae genomes in the different compartments. This information was summarized using the Venn Diagram package [81] with community indexes visualized as a dendrogram using the Vegan package [82] created in R v3.3.2 [83]. 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 Microviridae genome was not considered either in this study or in comparing these genomes to the dsDNA phages [55], since there are known biases created by RCA that lead to overrepresentation of ssDNA circular genomes [34,84,85].

3. Results

3.1. Diversity of Microviridae 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 members of the Microviridae. After clustering these genomes based on 95% genome-wide pairwise identity, a total of 258 genomes were identified and named with the prefix, Ciona gut microphage (CGM), followed by a number (Supplementary Table S2). 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 family Microviridae [65,86]. Coverage of the genomes in the compartment from which they were originally assembled ranged from 2× to 1167×, with a mean coverage of 35× (Supplementary Table S2).

To assess diversity, each CGM genome was annotated and MCP amino acid sequences were used for phylogenetic analysis (Figure 1). Based on this analysis, the vast majority (n = 188) of CGM sequences grouped with the established subfamily Gokushovirinae [87], followed by sequences closely related to the proposed Group D microviruses (n = 33) [69] and the proposed subfamily Pichovirinae (n = 20) [73]. A smaller proportion of CGM sequences were related to Parabacteroides prophages (n = 7) [69,88], and members of the proposed Alpavirinae (n = 2) subfamily [36,37]. Six of the Ciona gut MCP sequences (CGM_251, CGM_223, CGM_252, CGM_222, CGM_249, CGM_250) form a distinct group from all previously described sequences and are referred to here as “Novel CGM Group”. None of the CGM MCP sequences grouped with the established Bullavirinae subfamily [87] or the proposed Aravirinae [68], Stokavirinae [68], or Sukshmavirinae [41] groups. However, two CGM sequences were most closely related to the pequeñoviruses, a proposed sister clade of the Bullavirinae [69]. Since 73% of the CGM genomes group within the Gokushovirinae subfamily, this subfamily is presented in a separate tree (Figure 2). Almost half (47%, n = 88) of the CGM MCP sequences within the Gokushovirinae subfamily do not group with previously reported sequences and share less than 70% aa identity with known Gokushovirinae MCP sequences.

The gene synteny was compared between CGM genomes and previously reported Microviridae to evaluate if CGM genomes possess novel genome organizations (Figure 3). The CGM genomes expand upon previously identified gene organizations, yet many are syntenous with representative published sequences. The most diverse group, in terms of genome organization, was Group D, with 11 different patterns of gene organization.
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 proposed subfamilies (\( n = 96 \)). The tree was created using PhyML with aLRT-probabilities; the scale bar represents the number of amino acid substitutions per site. 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: the Gokushovirinae sub-tree is displayed in Figure 2. Accession numbers for sequences used in this analysis are listed in Supplementary Tables S1 and S2.
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; the scale bar represents the number of amino acid substitutions per site. Branches with probability values less than 0.7 were collapsed. Values greater than 0.7 are indicated at nodes. Clades highlighted in purple represent those where CGM sequences do not group with any previously described MCP sequences. Note: Accession numbers for sequences used in this analysis are listed in Supplementary Tables S1 and S2.
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 numbers (the 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. The novel CGM group does not have representatives in the database, while the Aravirinae, Stokavirinae, Sukshmavirinae, and Bullavirinae were not detected among the CGM sequences. Details on the gene order for each CGM genome are available in Supplementary Table S2.

3.2. Structure of Microviridae Communities

Richness of Microviridae genomes (i.e., the number of genomes that are <95% identical to each other) 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 genomes within the gut; these two compartments also share the largest number of genomes (Figure 4). Both water samples (MB & CB) group separately from the gut compartments (Figure 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 S2). Both the full and clear hindgut samples (HF and HC) 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 to that compartment. Interestingly, despite being full of dietary material, the full gut compartments have a lower overall richness than the cleared ones. Four of the six sequences (CGM_251, CGM_223, CGM_250, CGM_257) belonging to the novel CGM group were only found in the hindgut full (Supplementary Table S2), and not seen in any other gut compartments. MF had 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 to SF.

**Figure 4.** (A) Cluster dendrogram showing the relatedness among the CGM communities in the *Ciona* gut compartments and the surrounding water. (B) Venn diagrams showing comparisons between the closest groups on the dendrogram. (C) The three-way Venn diagrams specify shared and unique genomes detected in each of the compared groups. All diagrams were created based on the presence/absence of CGM genomes alone. The dendrogram was created using the Bray-Curtis dissimilarity index and the scale bar represents the dissimilarity values. Details on which CGM genomes were found in each compartment or water sample are available in Supplementary Table S2.
4. Discussion

Most microbiome research to date has focused on bacterial communities, with descriptions of the virome only recently gaining traction [25,27,89]. 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 [18,26,90]. Here, we described ssDNA phages found in the Ciona gut in an effort to further characterize the virome of this invertebrate model organism.

Recent applications of RCA in virome studies has dramatically increased our discovery of small, ssDNA viruses including phages belonging to the family *Microviridae*. These phages have now been described in a variety of habitats, including non-human animal guts [40–42,44–46,71,74,91], human guts [6,7,12,27,33,35,38,49,92–94], reclaimed water [95], sewage [96,97], fresh water systems [73,85,98–100], marine systems [24,100–108], methane seeps [70], modern stromatolites [109], confined aquifers [110], sediments [76,85,108,111–113], dragonflies [72], and fruit trees [114]. Despite this rapid increase in sequence information for ssDNA phages, their identification in such diverse environments has yet to reveal information about their hosts or functions. Now that a diversity of phages related to the *Microviridae* has been identified in Ciona, future experiments using this model organism can aim to define the role of these viruses in an animal gut.

In characterizing the diversity of *Microviridae*, we found 258 viral genomes within the Ciona gut, vastly outnumbering the total number of *Microviridae* genomes reported from any single animal gut study [6,33,35,37,41,42,74,94]. Although previous studies frequently identify sequences related to the *Microviridae* (most often belonging to the subfamily Gokushovirinae), detailed analysis of the genomes is rarely reported and contig sequences are often not available. Based on analysis of the MCP, which is typically used as a phylogenetic marker for this phage group [37,68,96,109], we found a variety of novel *Microviridae* groups. The CGM MCP sequence diversity encompassed 6 of the 10 established or proposed *Microviridae* groups in the literature, and a novel CGM clade was identified (Figure 1). Sequences belonging to the Gokushovirinae dominated in the Ciona gut, with 47% of the CGM gokushovirus MCP sequences forming clades without representatives from the literature (Figure 2). This finding adds to the growing body of literature stressing the predominance of gokushoviruses throughout numerous environments [12,33,35,37,70,98,105,107].

Interestingly, the only definitive hosts for Gokushovirinae are obligate parasitic or predatory bacteria [65] and laboratory isolates infect Bdellovibrio, Chlamydia, and Spiroplasma [86]. Although the hosts for the gokushoviruses identified in this study are not known, it should be noted that 16S ribosomal RNA gene data from a previous study of the Ciona gut [20] found representatives from each of these three known bacterial host groups. In addition, it is possible that gokushoviruses infect other types of bacteria with intracellular life stages. Other potential hosts for the gokushoviruses are the symbiotic bacteria belonging to the genus Endozoicomonas, which are dominant members of the core Ciona microbiome [20] and commonly associated with tunicates and other marine invertebrates from which gokushoviruses are frequently detected [24,115,116]. Finally, it is notable that Rickettsia were prevalent in the animals processed for this study [55], serving as another possible obligate intracellular bacterial host for the identified gokushoviruses.

Many of the CGMs were found to share a significant degree of synteny with previously observed members of the *Microviridae* (Figure 3). However, unique genome organizations were observed for several of the CGM genomes. The largest number of unique genome organizations was noted for Gokushovirinae and Group D viruses. The genomic features distinguishing the two established *Microviridae* subfamilies (Bullavirinae and Gokushovirinae), namely the MCP and scaffolding proteins, were conserved [65,86]. Based on these features, the vast majority of the CGM genomes are consistent with the Gokushovirinae subfamily, with the exception being the two CGM genomes grouping with the Bullavirinae sister clade, Pequeñoovirus, which have an external scaffolding protein [70].

Surprisingly, the large diversity of *Microviridae* observed in the Ciona gut was mostly absent from the surrounding seawater. The MB and CB water samples only contained 2 and 19 ssDNA phage
Viruses 2018, 10, 404

...genomes, respectively, all of which were present in the Ciona gut. The CB water originates from the holding tanks where the animals are placed between field collection and shipping. Though the animals spend less than 8 h in these waters, they are passing water through their siphons, feeding, and releasing feces, which could potentially contribute to the increased number of CGM genomes seen in this sample compared to MB, where the animals were originally collected. It is possible that some Microviridae virions are too small to be captured on a 100 kDa filter; however, the more likely explanation for the lack of overlap between the water and animal gut samples is that these viruses are less prevalent in seawater than the Ciona guts, especially if their hosts are intracellular bacteria. Nearly all of the CGM genomes (237 viral genomes) were unique to the Ciona gut (Figure 4), which parallels previous findings that dsDNA phages from the Ciona gut were also significantly different from those in the water column [55]. The fact that gokushovirus richness was much higher in the Ciona gut than in surrounding marine environments suggests that Ciona could be concentrating hosts for these phages, providing a model for further studies into their host range and infection dynamics.

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 phages are dispersed. However, distinct viral signatures are still found among the stomach, midgut and hindgut compartments, which can inform us of similarity among these 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 phages 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 and full) (Figure 4). These findings suggest that the midgut is likely an intermediate reservoir of phages 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, if the clearing process 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 Microviridae richness diminishes towards 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. For example, laboratory feeding experiments performed with fluorescently tagged food particles and/or bacteria in the Dishaw lab have revealed food pellets exiting the animals within 45 to 60 min after feeding. 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 and Leigh et al. [55] hypothesized that an increased prevalence of temperate phages may be due to prophage induction caused by the stress of clearing. Although originally thought to be strictly lytic, Microviridae sequences have been discovered as prophages within the genomes of Bacteroidetes and Parabacteroides species common in the human mouth and gut [36,69]. Seven CGM genomes with MCPs most similar to the Parabacteroides prophages were identified, suggesting the possibility that temperate ssDNA phages exist in this system. Although these sequences are in the minority in regards to the overall CGM diversity, it supports prior evidence that lysogeny is common in the Ciona gut [55].

This single study of the Ciona gut revealed more complete Microviridae genomes than any environmental [68–70,96,98,104–107] or gut study [6,27,33,37–39,41,42,94,117] to date. Many prior studies examining animal guts have defined the structure of the viral community but have not probed the diversity of ssDNA phages present. Studies reporting ssDNA phage genomes have identified low diversity in comparison to the CGM communities. As an example, one study examining the feces of patients with coronary heart disease found 12 Gokushovirinae genomes and 2 Microviridae genomes that did not group with any known subfamily [38]. Another study focusing on the guts of termites found 12 Microviridae genomes, 2 of which were Gokushovirinae, 3 that did not group with any reference genomes,
and 7 that they proposed to belong to a new subfamily Sukshmvirinae [41]. In comparison with prior studies, the remarkable Microviridae diversity (258 genomes) described in this study originate from the guts of only 10 Ciona individuals. These findings suggest that these phages, some of which may be infecting intracellular bacteria, likely infect hosts that are concentrated in or colonize 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 Ciona microbiome [20] can be manipulated and tightly controlled by rearing the animals germ-free [118]. Juvenile Ciona are small enough that dozens to hundreds of transparent juveniles can be reared on small tissue culture dishes, facilitating experimental manipulations. Therefore, the Ciona system affords many opportunities to address hypothesis-driven questions, possibly resulting in development of the first model for understanding the host range and biology of the Microviridae in animal guts.

Supplementary Materials: The following are available online at http://www.mdpi.com/1999-4915/10/8/404/s1, Table S1: Reference sequences used in constructing major capsid protein phylogenetic trees, Table S2: Details for each of the CGM phage genomes, including coverage, which compartment(s) they were detected in, and gene annotations.


Funding: This research was funded by National Science Foundation grant numbers IOS-1456301 and DEB-1555854. The APC was funded by National Science Foundation grant number IOS-1456301. A.C. was funded by the University of South Florida College of Marine Science’s Endowed Southern Kingfish Association Fellowship and George Lorton Fellowship in Marine Science. B.A.L. was funded by National Science Foundation Graduate Research Fellowship award number 1144244 and the University of South Florida College of Marine Science’s Endowed Mahaffey Family Graduate Fellowship in Marine Science.

Acknowledgments: The authors would like to thank Anni Djurhuus for assistance with R-code, Arvind Varsani for assistance with genome annotation, and Jonathan Vincent and Francois Enault for providing Microviridae MCP query sequences.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References


21. Sanders, J.G.; Beichman, A.C.; Roman, J.; Scott, J.J.; Emerson, D.; McCarthy, J.J.; Girguis, P.R. Baleen whales host a unique gut microbiome with similarities to both carnivores and herbivores. *Nat. Commun.* 2015, 6, 8285. [CrossRef] [PubMed]


41. Tikhe, C.V.; Husseneder, C. Metavirome sequencing of the termite gut reveals the presence of an unexplored bacteriophage community. *Front. Microbiol.* **2018**, *8*, 2548. [CrossRef] [PubMed]


60. Adriaenssens, E.; Brister, J.R. How to name and classify your phage: An informal guide. *Viruses*** **2018**, *10*, 404 15 of 18


79. Stöver, B.C.; Müller, K.F. TreeGraph 2: Combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinform.* 2010, 11, 7. [CrossRef] [PubMed]


© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).