

The RNA virome of echinoderms

Elliot W. Jackson^{1,2,*}, Roland C. Wilhelm³, Daniel H. Buckley^{1,3} and Ian Hewson¹

Abstract

Echinoderms are a phylum of marine invertebrates that include model organisms, keystone species, and animals commercially harvested for seafood. Despite their scientific, ecological, and economic importance, there is little known about the diversity of RNA viruses that infect echinoderms compared to other invertebrates. We screened over 900 transcriptomes and viral metagenomes to characterize the RNA virome of 38 echinoderm species from all five classes (Crinoidea, Holothuroidea, Asteroidea, Ophiuroidea and Echinoidea). We identified 347 viral genome fragments that were classified to genera and families within nine viral orders – *Picornavirales*, *Durnavirales*, *Martellivirales*, *Nodamuvirales*, *Reovirales*, *Amarillovirales*, *Ghabrivirales*, *Mononegavirales*, and *Hepelivirales*. We compared the relative viral representation across three life stages (embryo, larvae, adult) and characterized the gene content of contigs which encoded complete or near-complete genomes. The proportion of viral reads in a given transcriptome was not found to significantly differ between life stages though the majority of viral contigs were discovered from transcriptomes of adult tissue. This study illuminates the biodiversity of RNA viruses from echinoderms, revealing the occurrence of viral groups in natural populations.

INTRODUCTION

Metazoans harbour an enormous diversity and abundance of RNA viruses – a discovery that has reshaped our understanding of viral evolution through expanded viral-host associations, broadened phylogenetic diversity, and novel reconfigurations of genome architectures [1, 2]. Newly discovered viruses often blur the boundaries between well-known viral groups. For example, prior to a recent expansion, the family *Flaviviridae* was typified by relatively uniform, monopartite genomes, having a single 10–12 kb-long open reading frame (ORF), and being vectored to mammals by arthropods. Metagenomics has led to the discovery of hundreds of novel flavivirus genomes, redefining the genomic properties of this viral family and extending their host diversity beyond mammals [3–6]. To date, the exploration and systematization of invertebrate RNA viruses have been skewed towards terrestrial arthropods, mainly insects, leaving gaps in our understanding of the diversity, ecology, and evolution of RNA viruses in other invertebrate groups [1, 7–12]. To help close this gap, we characterized the RNA virome of Echinodermata – a phylum of marine invertebrates that are globally distributed throughout Earth's oceans and represent an evolutionary crossroads in developmental biology as one of two phyla that are invertebrate deuterostomes.

Wildlife disease and aquaculture are the two primary areas of concern regarding the threat of viral outbreaks among echinoderms. Disease outbreaks of sea urchins and sea stars have been documented at local, regional, and continental scales since 1898 and have gone unresolved in regards to their aetiology [13, 14]. Certain species of sea urchins and sea cucumbers are valued as seafood delicacies and the growing demands for these species in the seafood industry have led to a rise in aquaculture farming [15, 16]. Viral outbreaks pose a major concern for aquaculture operations [17, 18], yet little is known about the identity, let alone virulence, of viruses that infect these animals [19–21]. Baseline knowledge of viruses present in wild populations can help determine the aetiology of future outbreaks and discern pathogenic versus non-pathogenic agents, or those which may become more replicative under environmental stress. Thus, a census of viral diversity will improve our future response when viruses impact the economic or ecological function of echinoderms.

Parvoviruses – linear, single-stranded DNA viruses – are the best documented group known to infect echinoderms since the discovery of a densovirus in a sea urchin metagenome from Hawaii in 2014 [19]. Shortly after, another densovirus was found

Received 25 February 2022; Accepted 19 May 2022; Published 29 June 2022

Author affiliations: ¹Department of Microbiology, Cornell University, Ithaca, NY, USA; ²Scripps Institution of Oceanography, University of California San Diego, La Jolla CA, USA; ³School of Integrative Plant Science, Bradfield Hall, Cornell University, Ithaca, NY, USA.

***Correspondence:** Elliot W. Jackson, ewj34@cornell.edu

Keywords: viral discovery; Echinodermata; invertebrates; RNA viruses; transcriptome.

Three supplementary figures and two supplementary tables are available with the online version of this article.

in various sea star species and was implicated as the causative pathogen of the 2013/2014 Sea Star Wasting Syndrome (SSWS) outbreak in the Northeast Pacific [22]. However, subsequent attempts to correlate densoviruses to SSWS have not produced any clear association with pathology or disease [21–26]. Regardless, this discovery prompted a series of investigations into the diversity, prevalence, and association of these viruses with sea stars and SSWS [21, 24–26]. To date, no RNA virus identified using -omic approaches has been proven to cause any pathology in echinoderms [21, 27]. However, one clear line of evidence that has emerged from the accumulation of -omic data is that sea stars are infected by a diversity of viruses. We expect that they, and other echinoderms, will be host to a novel, undocumented diversity of RNA viruses.

Sequencing-based 'viromics' approaches have been the primary method for the discovery and characterization of echinoderm viruses. Other methods, like microscopy or culturing, are laborious and low throughput or hindered by the lack of available cell lines from aquatic invertebrates. All echinoderm virome studies, to date, have taken a viral metagenomic approach, where shotgun metagenomics is performed on encapsidated nucleic acids that have been enriched and selected for by chemical and/or nuclease treatment from size-filtered (<0.2 µm) tissue homogenates [19–22, 25–27]. Metatranscriptomes and transcriptomes are increasingly used for viral discovery and have only been applied to a few species of echinoderms [1, 28–32]. In the context of a metazoan, a metatranscriptome refers to the sequencing of total RNA (with rRNA removed) generally from samples pooled at the population level (multiple individuals of a species), while a transcriptome is the sequencing of poly-A tailed mRNA (with rRNA removed) generally from an individual organism [28, 33]. RNA-seq studies, for the purpose of viral discovery, typically generate or utilize metatranscriptomic datasets rather than transcriptomes, which are created for the purpose of analysing gene expression patterns of an organism. In this study, we analysed over 900 publicly available transcriptomes and viral metagenomes to characterize the biodiversity and distribution of RNA viruses associated with these animals. Together with previously published RNA viral metagenomes, we conducted a systematic survey of RNA viruses associated with the five major classes (Crinoidea, Holothuroidea, Asteroidea, Ophiuroidea and Echinoidea) of Echinodermata.

METHODS

The following sections detail the processing of the short-read libraries used for viral discovery and the analyses performed on the viral sequences. The two sources of libraries were transcriptomes and RNA-based viral metagenomes from various echinoderm species and tissues. All the viral metagenomes used in this study were generated by the authors. All libraries processed in this study were obtained from the NCBI's Sequence Read Archive (Table S1, available in the online version of this article). The transcriptome assemblies generated from this study are accessible through the Open Science Foundation (<https://osf.io/JXUAM> and <https://osf.io/BH8CR>).

Host transcriptomes and RNA-based viral metagenomes

A total of 903 paired-end transcriptomes derived from the five classes of echinoderm hosts, including crinoid ($n=18$; Crinoidea), sea cucumbers ($n=178$; Holothuroidea), sea star ($n=179$; Asteroidea), brittle star ($n=71$; Ophiuroidea) and urchin ($n=457$; Echinoidea). Raw sequences were quality controlled using Trimmomatic [34], to clip adapters, and FastX [35], to discard reads with lengths <50 nt and average quality scores <30. Transcriptomes were then assembled using default parameters in Trinity (v2.1.1) [36]. Contigs less than 500 nt were discarded prior to viral annotation.

A total of 24 paired-end RNA viral metagenomes derived from sea cucumbers ($n=3$; Holothuroidea) and sea stars ($n=21$; Asteroidea) were also retrieved from the NCBI's Sequence Read Archive (Table S1). Reads were merged then trimmed to remove adapters, and reads were discarded with average quality scores <20. Reads were then normalized to an average read depth of 100 using BBtools [37]. RNA viral metagenomes were assembled using Spades (v 3.11.1) with the -meta flag [38]. Contigs less than 500 nt were discarded prior to viral annotation.

Virus discovery and annotation

We curated a custom RNA virus database for viral annotation. The custom database contained viral amino acid sequences from invertebrates and aquatic environments [1, 12, 39] and from the NCBI viral genome database by selecting for viral sequences from invertebrates and invertebrates/vertebrates. Duplicated amino acid sequences were removed from the database using seqkit, yielding a total of 36 193 unique viral gene sequences. Echinoderm RNA viruses were then identified by querying transcriptome/RNA viral metagenome assemblies against the curated RNA viral database using DIAMOND BLASTx with the sensitivity parameter adjusted to 'very-sensitive' and an e-value cutoff of $<10^{-20}$ [40]. Contigs with significant similarity based on the BLAST criteria above were manually inspected in Geneious Prime (v 2020.2.2), and queried against the NCBI non-redundant database using the default BLASTp parameters to verify the viral annotation and assign putative gene function. BLASTp results were also used to identify conserved protein domains. By manually inspecting each viral contig we assessed genome completeness by expected genome length and complete coding regions of conserved genes. Genome illustrations were created by exporting the sequence viewer from Geneious Prime (v 2020.2.2) into Adobe Illustrator (v25.4.1). Contigs with near identical matches to human and plant viruses and bacteriophage were removed. ORFs containing an RdRP domain were used for taxonomic placement. Contigs

that did not contain a conserved protein domain were removed from further analysis. Quality-filtered reads were mapped to viral contigs to obtain relative abundance information using BBMap [41] using the 'semiperfect' flag which accommodated ambiguous bases (with equivalent results achieved with the 'perfect' flag).

Network analysis and phylogenetics

The taxonomic relationships of all recovered viral sequences were first mapped using a network analysis. To place sequences into broad taxonomic groups, we downloaded amino acid sequences from the top NCBI BLASTp results for each of our recovered viral sequences. A network was built based on sequence similarity using the online EFI-EST portal with default settings (minimum length=0, maximum length=50000, filter type=e-value $\leq 10^{-5}$ [42]. Nodes represent individual viral sequences and edges are the degree of similarity based off BLASTp pairwise similarity scores using a minimum pairwise similarity of 35%. Clusters and singletons were removed from the network that did not contain any representative viral sequences with a RdRP sequence. The network was visualized in Cytoscape (v 3.8.2) using the 'organic' layout [43].

We further established the relatives of echinoderm RNA viruses based on RdRP phylogenies. Independent phylogenetic analyses were performed for viral orders using the type species designated by the International Committee on Taxonomy of Viruses. RdRP amino acid sequences were aligned using MAFFT [44] and phylogenies were inferred by a substitution model selected by smart model selection in PhyML 3.0 with branch support determined by bootstrapping for 100 iterations [45]. The resulting phylogenetic tree was visualized and annotated using FigTree v1.4.4 [46]. The *Amarillovirales* phylogeny was created from the MAFFT alignment used in [47].

RESULTS

A total of 347 viral contigs, 33 of which represent complete or near-complete genomes, were recovered from the 927 short read libraries analysed. A total of 259 viral contigs were recovered from transcriptomes, and 88 viral contigs from RNA viral metagenomes (Figure S1B). The mean viral contig length recovered from RNA viral metagenomes (mean $\pm$ standard deviation: 4421 $\pm$ 3077 nt) was greater than from transcriptomes (3143 $\pm$ 3102 nt; Fig. S1a), and the size of sequencing libraries was weakly correlated with viral read depth in transcriptome libraries ($P=0.002$, Pearson's $r=0.29$) but not in RNA viral metagenomic libraries ($P=0.82$, $r=0.05$; Figure S1C). On average the relative abundance of viral contigs as a proportion of total reads was low, (0.0085%), ranging from 0.000023–0.29% ($\bar{x}=0.0085\%$). The average percentage of viral reads in viral metagenomes ($\bar{x}=0.39\%$) was ~45-fold higher than transcriptomes. However, viral abundance was highly uneven in the viral metagenomes with the average dropping to 0.07% (~eight-fold higher than transcriptomes) after excluding the top three most abundant samples.

Viral sequences were recovered from 111 of the 903 transcriptomic libraries and from all five echinoderm classes (Fig. 1). Sea cucumbers exhibited the highest prevalence of viral contigs among all echinoderm libraries (i.e. individuals) screened, followed by sea urchins (10%, 45/457), and the highest proportion among transcriptomes screened (26%, 47/178) (Fig. 1a). The majority of viral contigs recovered from transcriptomes came from adult tissue (70%) compared to embryos (20%) or larvae (10%) (Fig. 1B). In vertebrates, viral diversity and abundance has been found to be enriched in juveniles and younger individuals [48, 49]. Transcriptomes derived from echinoderms during their larval stage had a slightly higher proportion of viral reads in

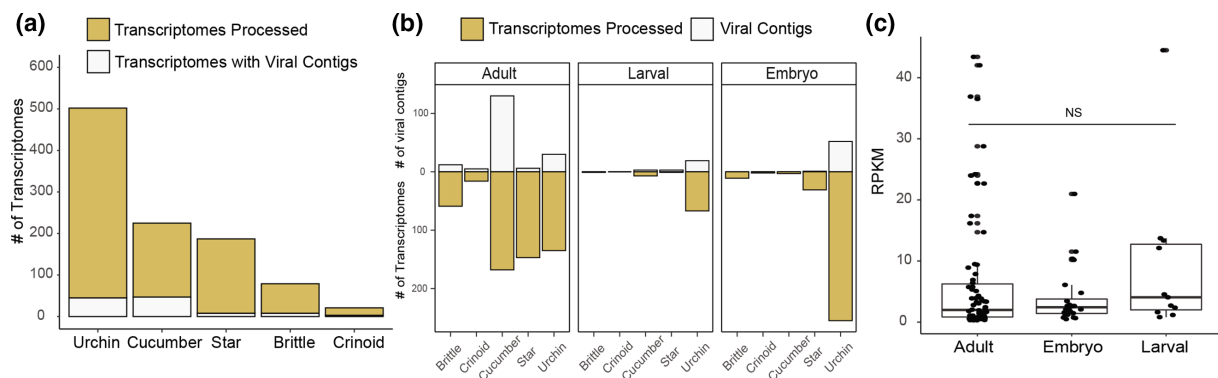


Fig. 1. Summary of viral contigs discovered from echinoderm transcriptomes (a) The number of transcriptomes downloaded from NCBI ordered by echinoderm class that were processed for viral discovery. (b) Top bars display the total number of viral contigs discovered separated by echinoderm class and life stage. Bottom bars display total number of transcriptomes separated by echinoderm class and life stage (c) Viral abundance in transcriptomes by life stage calculated by RPKM values. Kruskal-Wallis, $P>0.05$, NS=non-significant.

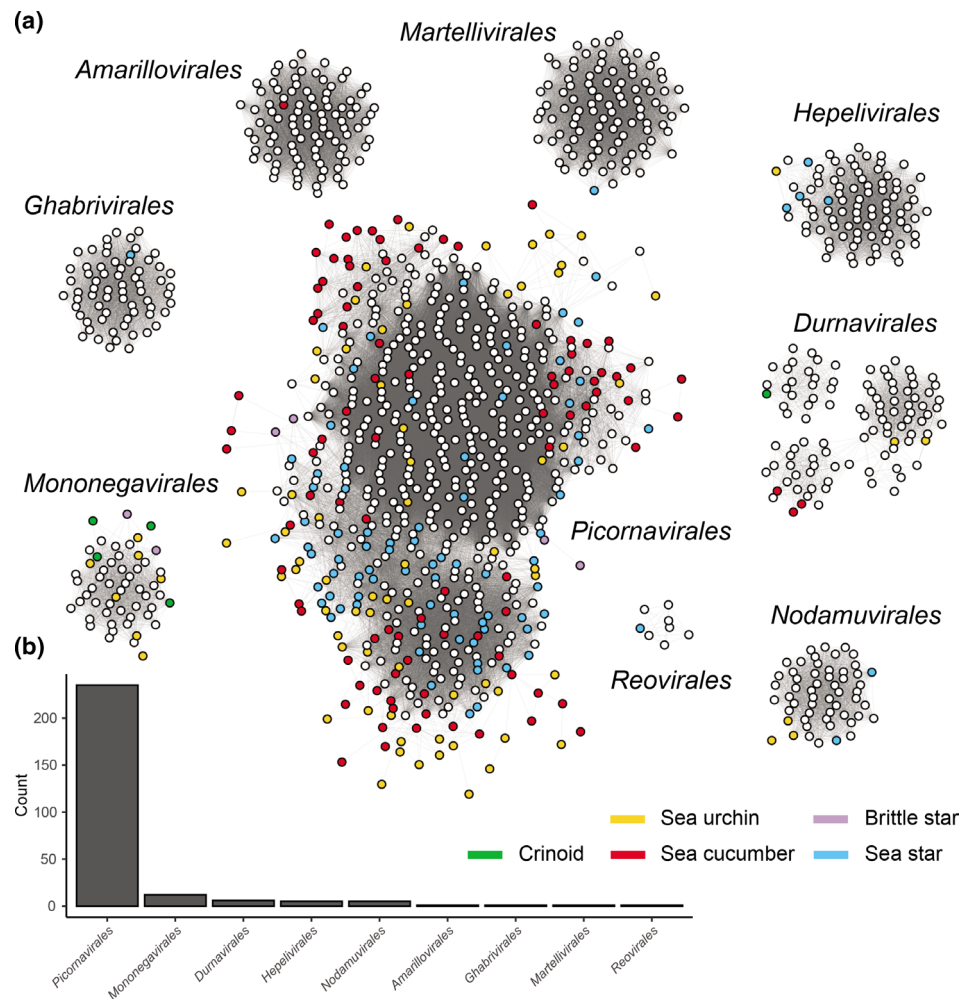


Fig. 2. Picornaviruses are the dominant viral order found in echinoderms. (a) Network analysis of viral sequences clustered by taxonomic order. Nodes represent individual viruses and edges are the degree of similarity based off BLASTp pairwise similarity using a minimum pairwise similarity of 35%. Coloured nodes represent viral sequences discovered from echinoderm transcriptomes and RNA viral metagenomes. White circles are viral genomes taken from NCBI. (b) The bar chart displays the number of echinoderm viruses discovered from each viral order.

their transcriptome ($\bar{x}$ =0.013%) than adults (0.0095%), and more than embryos (0.0034%; Fig. 1C), but these differences were not significant (Kruskal-Wallis, chi-squared 1.65, P -value=0.80).

Over half of the viral contigs contained an RdRP sequence (186/347), with 96 of these containing a complete or partial capsid sequence. Most viral contigs (215) contained at least a partial capsid sequence, and 42 viral contigs contained another conserved viral domain such as a methyltransferase or an RNA helicase domain (Table S2). The majority of viral contigs were taxonomically placed in the order *Picornavirales* (n =235) (Fig. 2). The recovered picornaviruses were distributed among a variety of families with the largest number related to *Marnaviridae*, followed by *Dicistroviridae*, and *Iflaviridae* (Fig. 3 and S2). Several unique clades within *Picornavirales* were represented by complete or near-complete genomes and may represent novel viral families (Fig. 2). The second highest number of viral contigs were recovered from *Mononegavirales* (n =12), with the remainder spread among seven orders (n =20): *Dumnavirales*, *Martellivirales*, *Nodamuvirales*, *Reovirales*, *Amarillovirales*, *Ghabrivirales*, and *Hepelivirales*. In general, the recovered viruses did not form new monophyletic clades within *Picornavirales* (Fig. 3 and S2) or *Amarillovirales*, *Reovirales*, and *Mononegavirales* (Fig. 4 and S3). However, within the *Hepelivirales*, the recovered viruses formed a distinct monophyletic clade that is sister to the clade containing *Orthohepivirus* and *Piscihepivirus*.

Viral contig lengths ranged from 502 nt to 12 989 nt. Complete and near complete genomes were recovered from *Picornavirales*, *Mononegavirales*, *Amarillovirales*, and *Hepelivirales* (Fig. 5). The echinoderm picornavirus genomes exhibited three different open reading frame arrangements, which spatially separated the genome by function according to replication or encapsidation (Fig. 3). The two most conserved protein domains related to replication were the RdRP (cd01699, pfam00978, pfam00946) and

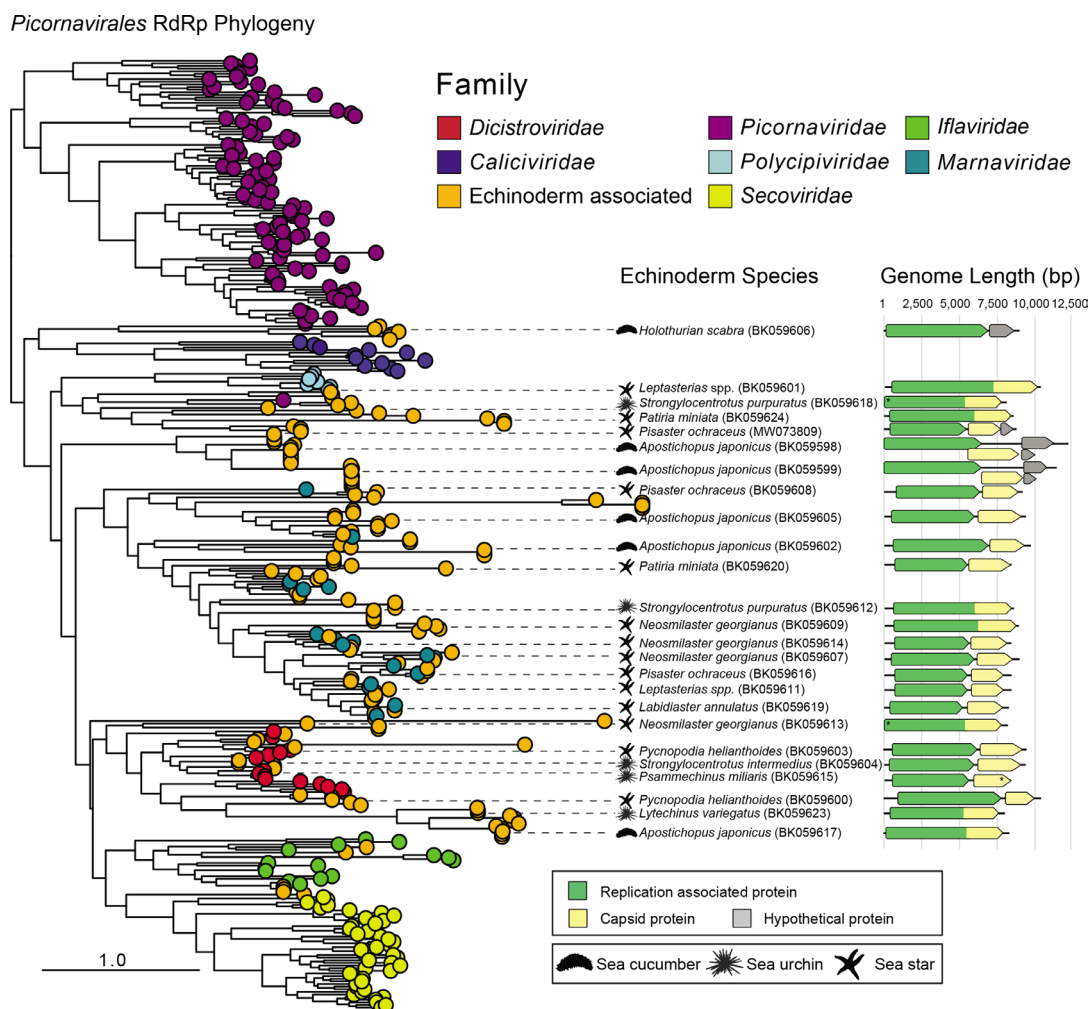


Fig. 3. Echinoderm picornaviruses are broadly distributed across the *Picornavirales* phylogeny. Tips are coloured by taxonomic family with orange circles representing echinoderm picornaviruses. Phylogenetic tree is rooted by *Picornaviridae*. Genome architectures of complete and near complete genomes recovered from assemblies displayed. Genomes are drawn approximate to scale in a 5' to 3' direction. Open reading frames denoted by boxes and coloured by general function. Asterisk represents an incomplete open reading frame. Animal icons represent the echinoderm order the viral contig is associated with.

RNA helicase (pfam00910, pfam01443) domains with many of the genomes also containing BIR (pfam00653), DSRM (cd00048), Sigma70 (pfam04539), peptidases (pfam12381), and large tegument protein (PHA03247) domains. The conserved capsid domains found among the picornavirus genome included: rhv-like (cd00205), dicistro VP4 (pfam11492), CRPV (pfam08762), and calici coat (pfam00915) domains (Fig. 5). The recovered hepeliviruses and mononegaviruses contigs contained the expected replication proteins but many lacked complete capsid proteins, indicating they were only near complete genomes (Fig. 5). The flavivirus genome previously discovered in sea cucumber [27] was completed during our assembly, leading to the extension of a second complete ORF and extending the genome size from 8883 to 12989 nt.

DISCUSSION

The most prevalent RNA viruses in echinoderm transcriptomes and RNA viral metagenomes were picornaviruses which are non-enveloped, single-stranded RNA (+ssRNA) viruses. The order *Picornavirales* is composed of eight families (*Caliciviridae*, *Dicistroviridae*, *Iflaviridae*, *Marnaviridae*, *Picornaviridae*, *Polycipiviridae*, *Secoviridae*, and *Soliniviridae*) and 103 genera and is among the most prevalent and diverse group of viruses found in -omics surveys of animal and environmental samples [30, 39, 50, 51]. The majority of the echinoderm-associated picornaviruses grouped into the *Dicistroviridae*, *Iflaviridae*, and *Marnaviridae* families, but other recovered viruses also comprised novel clades (Fig. 3 and S2) which supported our expectation that the extant diversity of echinoderm RNA viruses is under sampled. The *Marnaviridae* are known as ocean viroplankton which infect single-celled

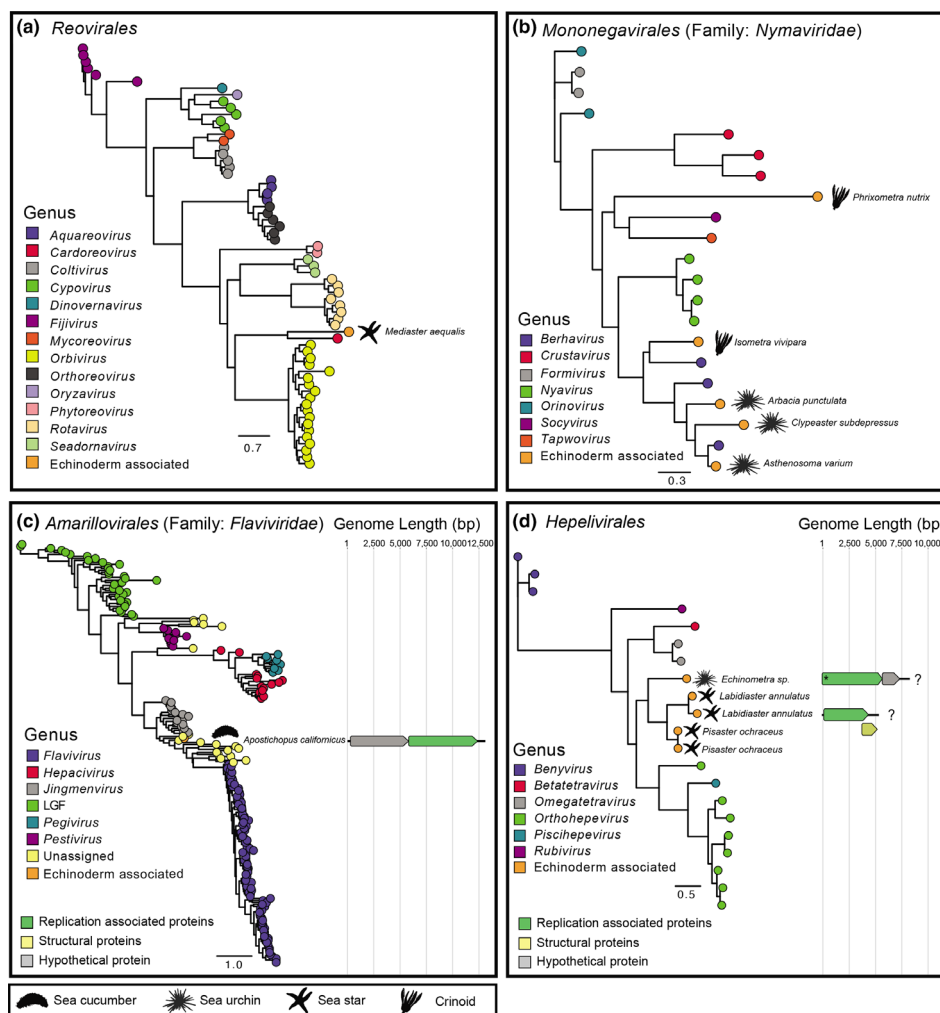


Fig. 4. Phylogenetic placement of echinoderm viruses from *Reovirales*, *Mononegavirales*, *Amarillovirales*, and *Hepelivirales*. Tips are coloured by taxonomic family or genus with black circles representing echinoderm viruses. Genome architectures of complete and near complete genomes recovered from assemblies displayed. Genomes are drawn approximate to scale in a 5' to 3' direction. Open reading frames denoted by boxes and coloured by general function. Asterisk represents an incomplete open reading frame and a question mark denotes potentially incomplete genome. Animal icons represent the echinoderm order the viral contig is associated with.

eukaryotes, such as phytoplankton and protists, but have also been found from metatranscriptomes from marine bivalves [51, 52]. It is possible that the *Marnaviridae* we observed infect protists that are symbionts or transiently associated with echinoderms. Alternatively, the host-range of the *Marnaviridae* family may extend beyond single-celled eukaryotes. The host range of many viral groups has changed considerably in recent years, and there are examples of host ranges within RNA viral families that do extend from protists to mammals, such as *Reoviridae* [53, 54]. All classified species of *Dicistroviridae* and *Iflaviridae* infect arthropods, and have largely been characterized due to the economic impacts of their pathogenicity though the host range of these families likely extends far beyond arthropods given their presence in metatranscriptomes from organisms in the phyla Mollusca, Cnidaria, and Platyhelminthes [32, 51, 55]. The disease severity from infections of both families ranges from inapparent to lethal, supporting the possibility that there may be non-pathogenic species infecting echinoderms.

Among the rare virosphere of echinoderms, and those that have few marine host associations, we observed *Martellivirales*, *Nodamuvirales*, *Reovirales*, *Amarillovirales*, *Mononegavirales*, and *Hepelivirales*. Many of these echinoderm viruses phylogenetically cluster with established invertebrate-infecting families or genera (i.e. *Nymaviridae* family within *Mononegavirales* or *Cardoreovirus* genus within *Reovirales*) while some represent evolutionary novel lineages such as the flavivirus discovered from a sea cucumber (Fig. 4C and S3c) [27]. Currently it is unclear if members of the rare virosphere infect all classes of echinoderms or if some are class specific. For example, the reovirus and flavivirus discovered in this study were only found in sea stars and sea cucumbers, respectively, though we cannot rule out methodological biases and insufficient sample size as proof of absence, requiring further

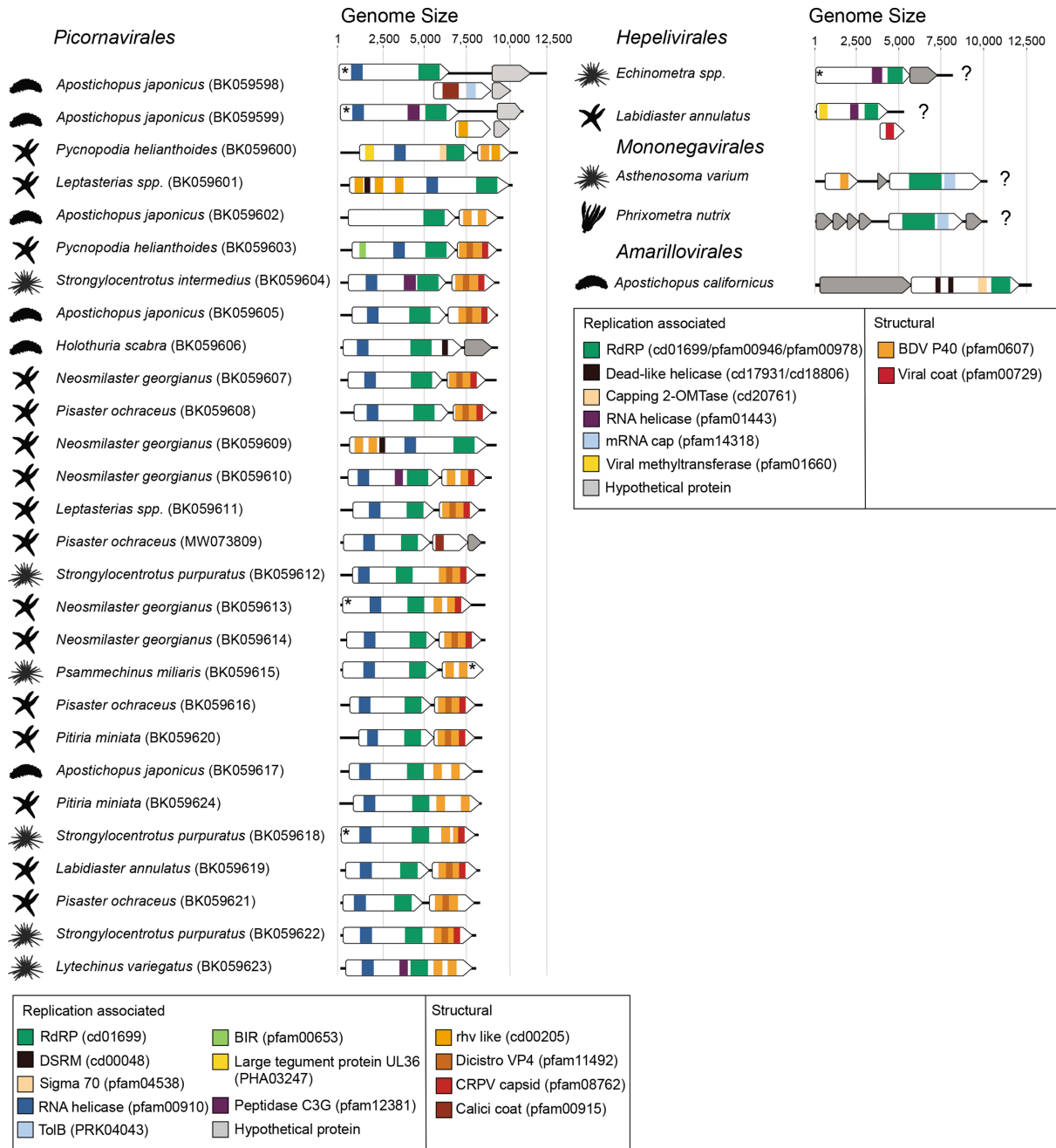


Fig. 5. Genome architectures and comparison of complete or near complete genomes recovered from assemblies. Genomes are drawn approximate to scale in a 5' to 3' direction. Open reading frames denoted by boxes and coloured regions represent conserved protein domains. Asterisk represents an incomplete open reading frame and question marks indicate missing open reading frames that would complete the genome. Animal icons represent the echinoderm order the viral contig is associated with.

research. Nevertheless, the discovery of these viruses significantly expands the host range for many of these groups and represents the first RNA viruses discovered from crinoids and brittle stars.

The vast majority of viruses recovered here, and elsewhere, using transcriptomic and metatranscriptomic approaches have been positive-sense single-stranded RNA (+ssRNA) [1, 51, 56]. This pattern of abundance likely has a basis in biology, but in the case of our dataset, may be inflated due to our use of transcriptomic data. The selection for polyadenylated transcripts during RNA-seq library preparation biases towards +ssRNA viruses, like picornaviruses, which have 3' polyadenylated tails as well as subgenomic RNA products that can be produced from various viral groups [57–59]. Studies utilizing a metatranscriptomic approach for RNA

viral discovery generally do not find such a highly skewed distribution towards +ssRNA but are nevertheless the most present viral type [1, 30, 39, 51]. By utilizing RNA viral metagenomes and transcriptomes we have uncovered the fullest diversity of RNA viruses associated with echinoderms with the datasets available though we expect future multi-omic efforts to reveal additional diversity.

The greatest difference between the two -omic approaches used in this study for viral discovery was the total number of viral contigs recovered and contig lengths. RNA viral metagenomes generally contained >3 viral contigs per library, which were ~40% longer, compared to transcriptomes, which contained 2.2 contigs per library. These differences are however likely driven by the assembler used, Trinity vs metaSPAdes, rather than underlying differences in the data. Additionally, the proportion of viral reads recovered exhibited a weak correlation with library size for transcriptomes but not for RNA viral metagenomes (Figure S1B). Thus, despite the efforts to enrich for viruses in the RNA viral metagenomes, the majority of libraries had a similar percent of viral reads compared to transcriptomes. These findings indicate that the efficacy of recovering RNA virus improves with sequencing depth, likely due to the improved assembly of sequences found in low abundance. The unbiased selection of RNA molecules in metatranscriptomes is often leveraged to measure viral sequence representation, referred to as abundance, that has led to an interesting observation that RNA viral abundance is typically greater in invertebrates compared to vertebrates [33]. The viral abundance found in the echinoderm transcriptomes is significantly lower than what has been reported for invertebrates, particularly Arthropods, but is likely driven by the biased selection for poly(A) tailed RNA. Future studies utilizing a metatranscriptome approach in echinoderms should reevaluate this observation to determine if viral abundance is similar to other invertebrate groups.

The capacity for viral discovery using -omic approaches has greatly expanded our understanding of biodiversity and host range, fundamentally shifting the perception of viruses as solely pathogens to a more nuanced role as commensals or mutualists [1]. Performing a viral census of hosts, like echinoderms, provides a useful context about the prevalence and association of viruses that can help understand future outbreaks or changes in the susceptibility of marine animals due to stress from climate change and human activity. The full potential of -omic approaches to understand the biological or ecological role of the diversity of viruses uncovered will only be fully realized in partnership with advances in culturing techniques to study the infection of naïve specimens [60, 61]. Our study provides a comprehensive survey of RNA viruses present in echinoderm, contrasting the diversity and abundance of RNA viruses between echinoderm classes and life stages. We hope this information provides valuable context for advancing our understanding of the role of these viruses in marine hosts and ecosystems.

Funding information

This work was supported by NSF grants OCE-1537111, OCE-1737127, and OCE-2049225 awarded to IH. This work was also supported by the Cornell Atkinson Centre's Sustainable Biodiversity Fund and Andrew W. Mellon Student Research Grant awarded to EWJ.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

- Shi M, Lin X-D, Tian J-H, Chen L-J, Chen X, et al. Redefining the invertebrate RNA virosphere. *Nature* 2016;540:539–543.
- Shi M, Lin X-D, Chen X, Tian J-H, Chen L-J, et al. The evolutionary history of vertebrate RNA viruses. *Nature* 2018;556:197–202.
- Bekal S, Domier LL, Gonfa B, McCoppin NK, Lambert KN, et al. A novel flavivirus in the soybean cyst nematode. *J Gen Virol* 2014;95:1272–1280.
- Fauver JR, Grubaugh ND, Krajacich BJ, Weger-Lucarelli J, Lakin SM, et al. West African Anopheles gambiae mosquitoes harbor a taxonomically diverse virome including new insect-specific flaviviruses, mononegaviruses, and totiviruses. *Virology* 2016;498:288–299.
- Qin X-C, Shi M, Tian J-H, Lin X-D, Gao D-Y, et al. A tick-borne segmented RNA virus contains genome segments derived from unsegmented viral ancestors. *Proc Natl Acad Sci U S A* 2014;111:6744–6749.
- Shi M, Lin X-D, Vasilakis N, Tian J-H, Li C-X, et al. Divergent viruses discovered in arthropods and vertebrates revise the evolutionary history of the *Flaviviridae* and related viruses. *J Virol* 2016;90:659–669.
- Junglen S, Drosten C. Virus discovery and recent insights into virus diversity in arthropods. *Curr Opin Microbiol* 2013;16:507–513.
- Li C-X, Shi M, Tian J-H, Lin X-D, Kang Y-J, et al. Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. *Elife* 2015;4.
- Liu S, Chen Y, Bonning BC. RNA virus discovery in insects. *Curr Opin Insect Sci* 2015;8:54–61.
- Remnant EJ, Shi M, Buchmann G, Blacqui re T, Holmes EC, et al. A Diverse Range of Novel RNA Viruses in Geographically Distinct Honey Bee Populations. *J Virol* 2017;91:e00158–17.
- K fer S, Paraskevopoulou S, Zirkel F, Wieseke N, Donath A, et al. Re-assessing the diversity of negative strand RNA viruses in insects. *PLoS Pathog* 2019;15:e1008224.
- Wu H, Pang R, Cheng T, Xue L, Zeng H, et al. Abundant and Diverse RNA Viruses in Insects Revealed by RNA-Seq Analysis: Ecological and Evolutionary Implications. *mSystems* 2020;5:e00039–20.
- Wang YN, Chang YQ, Lawrence JM. Chapter 12 - Disease in Sea Urchins. In: Lawrence JM (eds). *Developments in Aquaculture and Fisheries Science*. Elsevier. pp. 179–186.
- Hewson I, Sullivan B, Jackson EW, Xu Q, Long H, et al. Perspective: something old, something new? Review of wasting and other mortality in Asteroidea (Echinodermata). *Front Mar Sci* 2019;6:406.
- Slater M. Use and Exploitation of Sea Cucumbers. In: *Echinoderm Aquaculture*. John Wiley & Sons, Ltd. , pp. 57–73.
- Brown NP, Eddy SD. *Use and Exploitation of Sea Urchins*. Hoboken, NJ, USA: Echinoderm Aquaculture. John Wiley & Sons, Ltd; , pp. 25–45.
- Deng H, Zhou Z, Wang N, Liu C. The syndrome of sea cucumber (*Apostichopus japonicus*) infected by virus and bacteria. *Viral Sin* 2008;23:63–67.
- Liu H, Zheng F, Sun X, Hong X, Dong S, et al. Identification of the pathogens associated with skin ulceration and peristome tumescence in cultured sea cucumbers *Apostichopus japonicus* (Selenka). *J Invertebr Pathol* 2010;105:236–242.

19. Gudenkauf BM, Eaglesham JB, Aragundi WM, Hewson I. Discovery of urchin-associated densoviruses (family Parvoviridae) in coastal waters of the Big Island, Hawaii. *J Gen Virol* 2014;95:652–658.
20. Jackson EW, Bistolas KSI, Button JB, Hewson I. Novel circular single-stranded DNA viruses among an asteroid, echinoid and holothurian (Phylum: Echinodermata). *PLoS One* 2016;11:e0166093.
21. Hewson I, Aquino CA, DeRito CM. Virome variation during sea star wasting disease progression in *Pisaster ochraceus* (Asteroidea, Echinodermata). *Viruses* 2020;12:E1332.
22. Hewson I, Button JB, Gudenkauf BM, Miner B, Newton AL, et al. Densovirus associated with sea-star wasting disease and mass mortality. *Proc Natl Acad Sci U S A* 2014;111:17278–17283.
23. Bucci C, Francoeur M, McGreal J, Smolowitz R, Zazueta-Novoa V, et al. Sea star wasting disease in *Asterias forbesi* along the Atlantic Coast of North America. *PLoS One* 2017;12:e0188523.
24. Hewson I, Bistolas KSI, Quijano Cardé EM, Button JB, Foster PJ, et al. Investigating the complex association between viral ecology, environment, and Northeast Pacific Sea Star Wasting. *Front Mar Sci* 2018;5:77.
25. Jackson EW, Pepe-Rannek C, Johnson MR, Distel DL, Hewson I. A highly prevalent and pervasive densovirus discovered among sea stars from the North American Atlantic Coast. *Appl Environ Microbiol* 2020;86:e02723–19.
26. Jackson EW, Wilhelm RC, Johnson MR, Lutz HL, Danforth I, et al. Diversity of sea star-associated densoviruses and transcribed endogenous viral Elements of densovirus origin. *J Virol* 2020;95:e01594–20.
27. Hewson I, Johnson MR, Tibbetts IR. An unconventional *Flavivirus* and other RNA viruses in the sea cucumber (Holothuroidea; Echinodermata) Virome. *Viruses* 2020;12:E1057.
28. Shi M, Zhang Y-Z, Holmes EC. Meta-transcriptomics and the evolutionary biology of RNA viruses. *Virus Res* 2018;243:83–90.
29. Waldron FM, Stone GN, Obbard DJ. Metagenomic sequencing suggests a diversity of RNA interference-like responses to viruses across multicellular eukaryotes. *PLoS Genet* 2018;14:e1007533.
30. Wu H, Pang R, Cheng T, Xue L, Zeng H, et al. Abundant and diverse RNA viruses in insects revealed by RNA-Seq analysis: ecological and evolutionary implications. *mSystems* 2020;5:e00039–20.
31. Parry R, Asgari S, Williams BRG. Discovery of novel crustacean and cephalopod *Flaviviruses*: insights into the evolution and circulation of *Flaviviruses* between marine invertebrate and vertebrate hosts. *J Virol* 2019;93:e00432–19.
32. Cheng R-L, Li X-F, Zhang C-X. Novel dicistroviruses in an unexpected wide range of invertebrates. *Food Environ Virol* 2021;13:423–431.
33. Zhang Y-Z, Chen Y-M, Wang W, Qin X-C, Holmes EC. Expanding the RNA virosphere by unbiased metagenomics. *Annu Rev Virol* 2019;6:119–139.
34. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 2014;30:2114–2120.
35. Gordon A, Fastx-toolkit HGJ. FASTQ/A short-reads preprocessing tools; 2010. http://hannonlab.cshl.edu/fastx_toolkit
36. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 2011;29:644–652.
37. Bushnell B. BBTools software package; <http://sourceforge.net/projects/bbmap> 2014;578:579
38. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19:455–477.
39. Wolf YI, Silas S, Wang Y, Wu S, Bocek M, et al. Doubling of the known set of RNA viruses by metagenomic analysis of an aquatic virome. *Nat Microbiol* 2020;5:1262–1270.
40. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 2015;12:59–60.
41. Bushnell B. BMap: A Fast, Accurate, Splice-Aware Aligner.
42. Gerlt JA, Bouvier JT, Davidson DB, Imker HJ, Sadkhin B, et al. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochim Biophys Acta* 2015;1854:1019–1037.
43. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498–2504.
44. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;30:772–780.
45. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010;59:307–321.
46. Rambaut A, Drummond A. FigTree: Tree Figure Drawing Tool, Version 1.2.2. Institute of Evolutionary Biology, University of Edinburgh.
47. Paraskevopoulou S, Käfer S, Zirkel F, Donath A, Petersen M, et al. Viromics of extant insect orders unveil the evolution of the flavivirus superfamily. *Virus Evol* 2021;7:veab030.
48. Bergner LM, Orton RJ, Benavides JA, Becker DJ, Tello C, et al. Demographic and environmental drivers of metagenomic viral diversity in vampire bats. *Mol Ecol* 2020;29:26–39.
49. Wille M, Shi M, Hurt AC, Klaassen M, Holmes EC. RNA virome abundance and diversity is associated with host age in a bird species. *Virology* 2021;561:98–106.
50. Culley AI, Lang AS, Suttle CA. High diversity of unknown picorna-like viruses in the sea. *Nature* 2003;424:1054–1057.
51. Zhang Y-Y, Chen Y, Wei X, Cui J. Viromes in marine ecosystems reveal remarkable invertebrate RNA virus diversity. *Sci China Life Sci* 2022;65:426–437.
52. Lang AS, Vlok M, Culley AI, Suttle CA, Takao Y, et al. ICTV virus taxonomy profile: Marnaviridae 2021. *J Gen Virol* 2021;102:001633.
53. Brussaard CPD, Noordeeloos AAM, Sandaa R-A, Haldal M, Bratbak G. Discovery of a dsRNA virus infecting the marine photosynthetic protist *Micromonas pusilla*. *Virology* 2004;319:280–291.
54. Day JM. The diversity of the orthoreoviruses: molecular taxonomy and phylogenetic divides. *Infect Genet Evol* 2009;9:390–400.
55. Valles SM, Chen Y, Firth AE, Guérin DMA, Hashimoto Y, et al. ICTV virus taxonomy profile: Iflaviridae. *J Gen Virol* 2017;98:527–528.
56. Guo L, Lu X, Liu X, Li P, Wu J, et al. Meta-transcriptomic analysis reveals the virome and viral genomic evolution of medically important mites. *J Virol* 2020;JVI.01686–20.
57. Steil BP, Kempf BJ, Barton DJ. Poly(A) at the 3' end of positive-strand RNA and VPg-linked poly(U) at the 5' end of negative-strand RNA are reciprocal templates during replication of poliovirus RNA. *J Virol* 2010;84:2843–2858.
58. Visser M, Bester R, Burger JT, Maree HJ. Next-generation sequencing for virus detection: covering all the bases. *Virol J* 2016;13:1–6.
59. Gann ER, Kang Y, Dyhrman ST, Gobler CJ, Wilhelm SW. Metatranscriptome library preparation influences analyses of viral community activity during a brown tide bloom. *Front Microbiol* 2021;12:664189.
60. Canuti M, van der Hoek L. Virus discovery: are we scientists or genome collectors? *Trends Microbiol* 2014;22:229–231.
61. Greninger AL. A decade of RNA virus metagenomics is (not) enough. *Virus Res* 2018;244:218–229.