The RNA Virome of Echinoderms

Elliot W. Jackson[1,2*], Roland C. Wilhelm[3], Daniel H. Buckley[1,3], and Ian Hewson[1]

[1] Department of Microbiology, Cornell University, Ithaca NY USA

[2] Scripps Institution of Oceanography, University of California San Diego, La Jolla CA, USA

[3] School of Integrative Plant Science, Bradfield Hall, Cornell University, Ithaca, NY USA

 * Correspondence: ewj34@cornell.edu; Tel.: +1-231-838-6042

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

30

31   Metazoans harbor an enormous diversity and abundance of RNA viruses – a discovery

32   that has reshaped our understanding of viral evolution through expanded viral-host associations,

33   broadened phylogenetic diversity, and novel reconfigurations of genome architectures [1, 2].

34   Newly discovered viruses often blur the boundaries between well-known viral groups. For

35   example, prior to a recent expansion, the family *Flaviviridae* was typified by relatively uniform,

36   monopartite genomes, having a single 10-12 kb-long open reading frame (ORF), and being

37   vectored to mammals by arthropod. Metagenomics has led to the discovery of hundreds of novel

38   flavivirus genomes, redefining the genomic properties of this viral family, and extending their

39   host diversity beyond mammals [3–6]. To date, the exploration and systematization of

40   invertebrate RNA viruses have been skewed towards terrestrial arthropods, mainly insects,

41   leaving gaps in our understanding of the diversity, ecology, and evolution of RNA virus in other

42   invertebrate groups [1, 7–12]. To help close this gap, we characterized the RNA virome of

43   Echinodermata - a phylum of marine invertebrates that are globally distributed throughout

44   Earth's oceans and represent an evolutionary crossroads in developmental biology as one of two

45   phyla that are invertebrate deuterostomes.

46   Wildlife disease and aquaculture are the two primary areas of concern regarding the

47   threat of viral outbreaks among echinoderms. Disease outbreaks of sea urchins and sea stars have

48   been documented at local, regional, and continental scales since 1898 and have gone unresolved

49   in regards to their etiology [13, 14]. Certain species of sea urchins and sea cucumbers are valued

50   as seafood delicacies and the growing demands for these species in the seafood industry have led

51   to a rise in aquaculture farming [15, 16]. Viral outbreaks pose a major concern for aquaculture

52   operations [17, 18], yet little is known about the identity, let alone virulence, of viruses that

53   infect these animals [19–21]. Baseline knowledge of viruses present in wild populations can help

54   determine the etiology of future outbreaks and discern pathogenic versus non-pathogenic agents,

55   or those which may become more replicative under environmental stress. Thus, a census of viral

56   diversity will improve our future response when viruses impact the economic or ecological

57   function of echinoderms.

58   Parvoviruses - linear, single-stranded DNA viruses - are the best documented group

59   known to infect echinoderms since the discovery of a densovirus in a sea urchin metagenome

60   from Hawaii in 2014 [19]. Shortly after, another densovirus was found in various sea star species

61   and was implicated as the causative pathogen of the 2013/2014 Sea Star Wasting Syndrome

62   (SSWS) outbreak in the Northeast Pacific [22]. However, subsequent attempts to correlate

63   densoviruses to SSWS have not produced any clear association with pathology or disease [21–

64   26]. Regardless, this discovery prompted a series of investigations into the diversity, prevalence,

65   and association of these viruses with sea stars and SSWS [21, 24–26]. To date, no RNA virus

66   identified using -omic approaches has been proven to cause any pathology in echinoderms [21,

67   27]. However, one clear line of evidence that has emerged from the accumulation of -omic data

68   is that sea stars are infected by a diversity of viruses. We expect that they, and other

69   echinoderms, will be host to a novel, undocumented diversity of RNA viruses.

70   Sequencing-based 'viromics' approaches have been the primary method for the discovery

71   and characterization of echinoderm viruses. Other methods, like microscopy or culturing, are

72   laborious and low throughput or hindered by the lack of available cell lines from aquatic

73   invertebrates. All echinoderm virome studies, to date, have taken a viral metagenomic approach,

74   where shotgun metagenomics is performed on encapsidated nucleic acids that have been

75   enriched and selected for by chemical and/or nuclease treatment from size-filtered (<0.2 um)

76   tissue homogenates [19–22, 25–27]. Metatranscriptomes and transcriptomes are increasingly

77   used for viral discovery and have not yet been applied to echinoderms [28–31]. In the context of

78   a metazoan, a metatranscriptome refers to the sequencing of total RNA (with rRNA removed)

79   generally from samples pooled at the population level (multiple individuals of a species), while a

80   transcriptome is the sequencing of poly-A tailed mRNA (with rRNA removed) generally from an

81   individual organism [28, 32]. RNA-seq studies generating metatranscriptomes are typically for

82   the purpose of viral discovery as opposed to transcriptomes, which are created for the purpose of

83   analyzing gene expression patterns of an organism. In this study, we analyzed over 900 publicly

84   available transcriptomes from echinoderms to characterize the biodiversity and distribution of

85   RNA viruses. Together with previously published RNA viral metagenomes (i.e. a population of

86   genomes of RNA viruses), we conducted a systematic survey of RNA viruses associated with the

87   five major classes (Crinoidea, Holothuroidea, Asteroidea, Ophiuroidea and Echinoidea) of

88   Echinodermata.

Methods

89   The following sections detail the processing of the short-read libraries used for viral

90   discovery and the analyses performed on the viral sequences. The two sources of libraries were

91   transcriptomes and RNA-based viral metagenomes derived from various echinoderm species and

92   tissues. All libraries processed in this study were obtained from the NCBI's Short Read Archive

93   (Table S1) and the assemblies generated from this study and the database used for viral

94   discovery are accessible through the Open Science Foundation (https://osf.io/JXUAM).

95   *Host transcriptomes and RNA-based viral metagenomes*

96

97   A total of 903 paired-end transcriptomes derived from the five classes of echinoderm

98   hosts, including crinoid (n = 18; Crinoidea), sea cucumbers (n = 178; Holothuroidea), sea star (n

99   = 179; Asteroidea), brittle star (n = 71; Ophiuroidea) and urchin (n = 457; Echinoidea). Raw

100   sequences were quality controlled using Trimmomatic [33], to clip adapters, and FastX [34], to

101   discard reads with lengths < 50 nt and average quality scores < 30. Transcriptomes were then

102   assembled using default parameters in Trinity (v2.1.1) [35]. Contigs less than 500 nt were

103   discarded prior to viral annotation.

104   A total of 24 paired-end RNA viral metagenomes derived from sea cucumbers (n=3;

105   Holothuroidea) and sea stars (n=21; Asteroidea) were also retrieved from the Short Read Archive

106   (Table S1). Raw sequences were merged, trimmed to clip adapters and discard reads with

107   average quality scores < 20, and normalized to an average read depth of 100 using BBtools [36].

108   RNA viral metagenomes were assembled using Spades (v 3.11.1) with the -meta flag [37].

109   Contigs less than 500 nt were discarded prior to viral annotation.

110   *Virus discovery and annotation*

111

112   We curated an RNA virus database for viral annotation. The database contained viral

113   amino acid sequences from Shi et al 2016, Wu et al 2020 and Wolf et al 2020, and from the

114   NCBI viral genome database after filtering for viral sequences from invertebrates and

115   invertebrates/vertebrates [1, 29, 38]. Duplicated amino acid sequences were removed from the

116   database using seqkit, yielding a total of 36,193 unique viral gene sequences. Echinoderm RNA

117   viruses were then identified by querying transcriptome/RNA viral metagenome assemblies

118   against the curated RNA viral database using DIAMOND BLASTx with the sensitivity

119   parameter adjusted to 'very-sensitive' and an e-value cutoff of $< 10^{-20}$ [39]. Contigs with

120   significant similarity based on the BLAST criteria above were manually inspected in Geneious

121   Prime (v 2020.2.2), and queried against the NCBI non-redundant database using the default

122   BLASTp parameters to verify the viral annotation and assign putative gene function. BLASTp

123   results were also used to identify conserved protein domains. Genome illustrations were created

124   by exporting the sequence viewer from Geneious Prime (v 2020.2.2) into Adobe Illustrator

125   (v25.4.1). Contigs with near identical matches to human and plant viruses and bacteriophage

126   were removed. ORFs containing an RdRP domain were used for taxonomic placement. If a

127   contig did not contain an RdRP sequence, a complete or partial ORF containing any conserved

128   protein domain was chosen. Contigs that did not contain a conserved protein domain were

129   removed from further analysis. Quality-filtered reads were mapped to viral contigs to obtain

130   relative abundance information using BBMap [40] using the 'semiperfect' flag which

131   accommodated ambiguous bases (with equivalent results achieved with the 'perfect' flag).

132   *Network analysis and phylogenetics*

133

134   The taxonomic relationships of all recovered viral sequences were first mapped using a

135   network analysis. To place sequences into broad taxonomic groups, we downloaded amino acid

136   sequences from the top NCBI BLASTp results for each of our recovered viral sequences. A

137   network was built based on sequence similarity using the online EFI-EST portal with default

138   settings (minimum length = 0, maximum length = 50,000, filter type = e-value $\leq 10^{-5}$ [41]. Nodes

139   represent individual viral sequences and edges are the degree of similarity based off BLASTp

140   pairwise similarity scores using a minimum pairwise similarity of 35%. Clusters and singletons

141   were removed from the network that did not contain any representative viral sequences with a

142   RdRP sequence. The network was visualized in Cytoscape (v 3.8.2) using the 'organic' layout

143   [42].

144   We further established the relatives of echinoderm RNA viruses based on RdRP

145   phylogenies. Independent phylogenetic analyses were performed for viral orders using the type

146   species designated by the International Committee on Taxonomy of Viruses. RdRP amino acid

147   sequences were aligned using MAFFT [43] and phylogenies were inferred by a substitution

148   model selected by smart model selection in PhyML 3.0 with branch support determined by

149   bootstrapping for 100 iterations [44]. The resulting phylogenetic tree was visualized and

150   annotated using FigTree v1.4.4 [45]. The *Amarillovirales* phylogeny was created from the

151   MAFFT alignment used in [46].

152   **Results**

153   We recovered a total of 347 viral contigs and 33 complete or near-complete genomes

154   from the 927 short read libraries analyzed. A total of 259 viral contigs were recovered from

155   transcriptomes, and 88 viral contigs from RNA viral metagenomes (Figure S1B). The mean viral

156   contig length recovered from RNA viral metagenomes (mean $\pm$ standard deviation: $4,421 \pm 3077$

157   nt) was greater than from transcriptomes ($3,143 \pm 3102$ nt; Figure S1A), and the size of

158   sequencing libraries was weakly correlated with viral read depth in transcriptome libraries ($p =$

159   0.002, Pearson's $r = 0.29$) but not in RNA viral metagenomic libraries ($p = 0.82$, $r = 0.05$; Figure

160   S1C). On average the relative abundance of viral contigs as a proportion of total reads was low,

161   (0.0085%), ranging from 0.000023% to 0.29% ($x\square = 0.0085\%$). The average percentage of viral

162   reads in viral metagenomes ($x\square = 0.39\%$) was ~45-fold higher than transcriptomes. However,

163   viral abundance was highly uneven in the viral metagenomes with the average dropping to

164   0.07% (~8-fold higher than transcriptomes) after excluding the top three most abundant samples.

165   Viral sequences were recovered from 111 of the 903 transcriptomic libraries and from all

166   five echinoderm classes (Figure 1). Sea cucumbers exhibited the highest prevalence of viral

167   contigs among all echinoderm libraries (*i.e.,* individuals) screened, followed by sea urchins

168   (10%, 45/457), and the highest proportion among transcriptomes screened (26%; 47/178) (Figure

169   1A). The majority of viral contigs recovered from transcriptomes came from adult tissue (70%)

170   compared to embryos (20%) or larvae (10%) (Figure 1B). Transcriptomes derived from

171   echinoderms during their larval stage had a slightly higher proportion of viral reads in their

172   transcriptome ($\bar{x}$ = 0.013%) than adults (0.0095%), and more than embryos (0.0034%; Figure

173   1C), but these differences were not significant (Kruskal-Wallis, chi-squared 1.65, p-value= 0.80).

174   Over half of the viral contigs contained an RdRP sequence (186/347), with 96 of these

175   containing a complete or partial capsid sequence. Most viral contigs (215) contained at least a

176   partial capsid sequence, and 42 viral contigs contained another conserved viral domain such as a

177   methyltransferase or an RNA helicase domain (Supplemental Table 2). The majority of viral

178   contigs were taxonomically placed in the order *Picornavirales* (n= 235) (Figure 2). The

179   recovered picornaviruses were distributed among a variety of families with the largest number

180   related to *Marnaviridae*, followed by *Dicistroviridae*, and *Iflaviridae* (Figure 3). Several unique

181   clades within *Picornavirales* were represented by complete or near-complete genomes and may

182   represent novel viral families (Figure 2). The second highest number of viral contigs were

183   recovered from *Mononegavirales* (n = 12)*,* with the remainder spread among seven orders (n=

184   20): *Durnavirales, Martellivirales, Nodamuvirales, Reovirales, Amarillovirales, Ghabrivirales,*

185   and *Hepelivirales*. In general, the recovered viruses did not form new monophyletic clades

186   within *Picornavirales* (Figure 3) or *Amarillovirales*, *Reovirales*, and *Mononegavirales* (Figure

187   4). However, within the *Hepivirales*, the recovered viruses formed a distinct monophyletic clade

188   that is sister to the clade containing *Orthohepivirus* and *Piscihepivirus*.

189   Viral contig lengths ranged from 502 nt to 12,989 nt. Complete and near complete

190   genomes were recovered from *Picornavirales, Mononegavirales, Amarillovirales,* and

191   *Hepelivirales* (Figure 5). The echinoderm picornavirus genomes exhibited three different open

192   reading frame arrangements, which spatially separated the genome by function according to

193   replication or encapsidation (Figure 3). The two most conserved protein domains related to

194   replication were the RdRP (pfam00680) and RNA helicase (pfam00910) domains with many of

195   the genomes also containing BIR (pfam00653), DSRM (pfam00035), Sigma70 (pfam04539),

196   peptidases (pfam12381), and large tegument protein (PHA03247) domains. The conserved

197   capsid domains found among the picornavirus genome included: rhv-like (pfam00073), dicistro

198   VP4 (pfam11492), CRPV (pfam08762), and calici coat (pfam00915) domains (Figure 5). The

199   recovered hepeviruses and mononegaviruses contigs contained the expected replication proteins

200   but many lacked complete capsid proteins, indicating they were only near complete genomes

201   (Figure 5). The flavivirus genome previously discovered in sea cucumber [27] was completed

202   during our assembly, leading to the extension of a second complete ORF and extending the

203   genome size from 8,883 to 12,989 nt.

204   **Discussion**

205   The most prevalent RNA viruses in echinoderm transcriptomes and RNA viral

206   metagenomes were picornaviruses which are non-enveloped, single-stranded RNA (+ssRNA)

207   viruses. The order *Picornavirales* is comprised of eight families (*Caliciviridae, Dicistroviridae,*

208   *Iflaviridae, Marnaviridae, Picornaviridae, Polycipiviridae, Secoviridae,* and *Solinviviridae*) and

209   103 genera and are among the most prevalent and diverse group of viruses found in -omics

210   surveys of animal and environmental samples [47, 29, 38, 48]. The majority of the echinoderm-

211   associated picornaviruses grouped into the *Dicistroviridae, Iflaviridae,* and *Marnaviridae*

212   families, but other recovered viruses also comprised novel clades (Figure 3) which supported our

213   expectation that the extant diversity of echinoderm RNA viruses is under sampled. The

214   *Marnaviridae* are known to be ocean virioplankton which infect single-celled eukaryotes, such

215   as phytoplankton and protists, but have also been found from metatranscriptomes from marine

216   bivalves [48, 49]. It is possible that the *Marnaviridae* we observed infect protists that are

217   symbionts or transiently associated with echinoderms. Alternatively, the host-range of the

218   *Marnaviridae* family may extend beyond single-celled eukaryotes. The host range of many viral

219   groups has changed considerably in recent years, and there are examples of host ranges within

220   RNA viral families that do extend from protists to mammals, such as *Reoviridae* [50, 51]. All

221   classified species of *Dicistroviridae* and *Iflaviridae* infect arthropods, and have largely been

222   characterized due to the economic impacts of their pathogenicity though the host range of these

223   families likely extends far beyond arthropods given their presence in metatranscriptomes from

224   organisms in the phyla Mollusca, Cnidaria, and Platyhelminthes [31, 48, 52]. The disease

225   severity from infections of both families ranges from inapparent to lethal, supporting the

226   possibility that there may be non-pathogenic species infecting echinoderms.

227   Among the rare virosphere of echinoderms, and those that have few marine host

228   associations, we observed *Martellivirales, Nodamuvirales, Reovirales, Amarillovirales,*

229   *Mononegavirales,* and *Hepelivirales*. Many of these echinoderm viruses phylogenetically cluster

230   with established invertebrate-infecting families or genera (i.e. *Nymaviridae* family within

231   *Mononegavirales* or *Cardorevirus* genus within *Reovirales*) while some represent evolutionary

232   novel lineages such as the flavivirus discovered from a sea cucumber (Figure 4C) [27]. Currently

233   it is unclear if members of the rare virosphere infect all classes of echinoderms or if some are

234   class specific. For example, the reovirus and flavivirus discovered in this study were only found

235   in sea stars and sea cucumbers, respectively, though we cannot rule out methodological biases

236   and insufficient sample size as proof of absence, requiring further research. Nevertheless, the

237   discovery of these viruses significantly expands the host range for many of these groups and

238   represents the first RNA viruses discovered from crinoids and brittle stars.

239   The vast majority of viruses recovered here, and elsewhere, using transcriptomic and

240   metatranscriptomic approaches have been positive-sense single-stranded RNA (+ssRNA) [1, 48,

241   53]. This pattern of abundance likely has a basis in biology, but in the case of our dataset, may be

242   inflated due to our use of transcriptomic data. The selection for polyadenylated transcripts during

243   RNA-seq library preparation biases towards +ssRNA viruses, like picornaviruses, which have

244   3'polyadenylated tails [54, 55]. Studies utilizing a metatranscriptomic approach for RNA viral

245   discovery generally do not find such a highly skewed distribution towards +ssRNA but are

246   nevertheless the most abundant viral type [1, 29, 38, 48]. By utilizing viral RNA metagenomes

247   and transcriptomes we have uncovered the fullest diversity of RNA viruses associated with

248   echinoderms with the datasets available, though we expect future multi-omic efforts to reveal

249   additional diversity.

250   The greatest difference between the two -omic approaches used in this study for viral

251   discovery was the total number of viral contigs recovered and contig lengths. Viral RNA

252   metagenomes generally contained >3 viral contigs per library, which were ~40 % longer,

253   compared to transcriptomes, which contained 2.2 contigs per library. Additionally, the

254   proportion of viral reads recovered exhibited a weak correlation with library size for

255   transcriptomes but not for viral RNA metagenomes (Figure S1B). Thus, despite the efforts to

256   enrich for viruses in the viral RNA metagenomes, the majority of libraries had a similar percent

257   of viral reads compared to transcriptomes. Furthermore, these findings indicate that the efficacy

258   of recovering RNA virus improves with sequencing depth, likely due to the improved assembly

259   of sequenced found in low abundance.

260   The capacity for viral discovery using -omic approaches has greatly expanded our

261   understanding of biodiversity and host range, fundamentally shifting the perception of viruses as

262   solely pathogens to a more nuanced role as commensals or mutualists [1]. Performing a viral

263   census of hosts, like echinoderms, provides a useful context about the prevalence and association

264   of viruses that can help understand future outbreaks or changes in the susceptibility of marine

265   animals due to stress from climate change and human activity. The full potential of -omic

266   approaches to understand the biological or ecological role of the diversity of viruses uncovered

267   will only be fully realized in partnership with advances in culturing techniques to study the

268   infection of naïve specimens [56, 57]. Our study provides a comprehensive survey of RNA

269   viruses present in echinoderm, contrasting the diversity and abundance of RNA viruses between

270   echinoderm classes and life stages. We hope this information provides valuable context for

271   advancing our understanding of the role of these viruses in marine hosts and ecosystems.

272   **Funding**

276   **Author statements**

277   The authors declare that there are no conflicts of interest.

278

References

1. **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.

2. **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.

3. **Bekal S, Domier LL, Gonfa B, McCoppin NK, Lambert KN, *et al.*** A novel flavivirus in the soybean cyst nematode. *Journal of General Virology* 2014;95:1272–1280.

4. **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.

5. **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. *Proceedings of the National Academy of Sciences* 2014;111:6744–6749.

6. **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. *Journal of Virology* 2016;90:659–669.

7. **Junglen S, Drosten C.** Virus discovery and recent insights into virus diversity in arthropods. *Current Opinion in Microbiology* 2013;16:507–513.

8. **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:e05378.

9. **Liu S, Chen Y, Bonning BC.** RNA virus discovery in insects. *Current Opinion in Insect Science* 2015;8:54–61.

10. **Remnant EJ, Shi M, Buchmann G, Blacquière T, Holmes EC, *et al.*** A diverse range of novel RNA viruses in geographically distinct honey bee populations. *Journal of virology* 2017;91:e00158-17.

11. **Käfer S, Paraskevopoulou S, Zirkel F, Wieseke N, Donath A, *et al.*** Re-assessing the diversity of negative strand RNA viruses in insects. *PLoS Pathogens* 2019;15:e1008224.

12. **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.

13. **Wang YN, Chang YQ, Lawrence JM.** Chapter 12 - Disease in Sea Urchins. In: Lawrence JM (editor). *Developments in Aquaculture and Fisheries Science*. Elsevier. pp. 179–186.

312 14. **Hewson I, Sullivan B, Jackson EW, Xu Q, Long H, *et al.*** Perspective: Something Old,
313   Something New? Review of Wasting and Other Mortality in Asteroidea (Echinodermata).
314   *Frontiers in Marine Science* 2019;6:406.

315 15. **Slater M**. Use and Exploitation of Sea Cucumbers. In: *Echinoderm Aquaculture*. John Wiley
316   & Sons, Ltd. pp. 57–73.

317 16. **Sun J, Chiang F-S**. Use and Exploitation of Sea Urchins. In: *Echinoderm Aquaculture*. John
318   Wiley & Sons, Ltd. pp. 25–45.

319 17. **Deng H, Zhou Z, Wang N, Liu C**. The syndrome of sea cucumber (Apostichopus japonicus)
320   infected by virus and bacteria. *Virologica Sinica* 2008;23:63–67.

321 18. **Liu H, Zheng F, Sun X, Hong X, Dong S, *et al.*** Identification of the pathogens associated with
322   skin ulceration and peristome tumescence in cultured sea cucumbers Apostichopus
323   japonicus (Selenka). *Journal of Invertebrate Pathology* 2010;105:236–242.

324 19. **Gudenkauf BM, Eaglesham JB, Aragundi WM, Hewson I**. Discovery of urchin-associated
325   densoviruses (family Parvoviridae) in coastal waters of the Big Island, Hawaii. *Journal of
326   General Virology* 2014;95:652–658.

327 20. **Jackson EW, Bistolas KS, Button JB, Hewson I**. Novel Circular Single-Stranded DNA Viruses
328   among an Asteroid, Echinoid and Holothurian (Phylum: Echinodermata). *PloS One*
329   2016;11:e0166093.

330 21. **Hewson I, Aquino CA, DeRito CM**. Virome variation during sea star wasting disease
331   progression in Pisaster ochraceus (Asteroidea, Echinodermata). *Viruses* 2020;12:1332.

332 22. **Hewson I, Button JB, Gudenkauf BM, Miner B, Newton AL, *et al.*** Densovirus associated
333   with sea-star wasting disease and mass mortality. *Proceedings of the National Academy of
334   Sciences* 2014;111:17278–17283.

335 23. **Bucci C, Francoeur M, McGreal J, Smolowitz R, Zazueta-Novoa V, *et al.*** Sea Star Wasting
336   Disease in Asterias forbesi along the Atlantic Coast of North America. *PLoS One*
337   2017;12:e0188523.

338 24. **Hewson I, Bistolas KS, Quijano Cardé EM, Button JB, Foster PJ, *et al.*** Investigating the
339   complex association between viral ecology, environment, and northeast Pacific sea star
340   wasting. *Frontiers in Marine Science* 2018;5:77.

341 25. **Jackson EW, Pepe-Ranney C, Johnson MR, Distel DL, Hewson I**. A Highly Prevalent and
342   Pervasive Densovirus Discovered among Sea Stars from the North American Atlantic Coast.
343   *Applied and Environmental Microbiology* 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. *Journal of Virology* 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;12:1057.

28. Shi M, Zhang Y-Z, Holmes EC. Meta-transcriptomics and the evolutionary biology of RNA viruses. *Virus Research* 2018;243:83–90.

29. 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*;5.

30. Parry R, Asgari S. Discovery of novel crustacean and cephalopod flaviviruses: insights into the evolution and circulation of flaviviruses between marine invertebrate and vertebrate hosts. *Journal of Virology* 2019;93:e00432-19.

31. Cheng R-L, Li X-F, Zhang C-X. Novel Dicistroviruses in an Unexpected Wide Range of Invertebrates. *Food and Environmental Virology* 2021;1–9.

32. Zhang Y-Z, Chen Y-M, Wang W, Qin X-C, Holmes EC. Expanding the RNA virosphere by unbiased metagenomics. *Annual Review of Virology* 2019;6:119–139.

33. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;30:2114–2120.

34. Gordon A, Hannon GJ. Fastx-toolkit. *FASTQ/A short-reads preprocessing tools* 2010;5:http://hannonlab.cshl.edu/fastx_toolkit.

35. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 2011;29:644–652.

36. Bushnell B. BBTools software package. *URL http://sourceforge net/projects/bbmap* 2014;578:579.

37. 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. *Journal of computational biology* 2012;19:455–477.

38. 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. *Nature Microbiology* 2020;5:1262–1270.

39. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nature Methods* 2014;12:59–60.

376 40. **Bushnell B**. BBMap: A Fast, Accurate, Splice-Aware Aligner.

377 41. **Gerlt JA, Bouvier JT, Davidson DB, Imker HJ, Sadkhin B, *et al.*** Enzyme function initiative-
378   enzyme similarity tool (EFI-EST): a web tool for generating protein sequence similarity
379   networks. *Biochimica Et Biophysica Acta (BBA)-Proteins and Proteomics* 2015;1854:1019–
380   1037.

381 42. **Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, *et al.*** Cytoscape: a software
382   environment for integrated models of biomolecular interaction networks. *Genome research*
383   2003;13:2498–2504.

384 43. **Katoh K, Standley DM**. MAFFT multiple sequence alignment software version 7:
385   improvements in performance and usability. *Molecular biology and evolution* 2013;30:772–
386   780.

387 44. **Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, *et al.*** New algorithms and
388   methods to estimate maximum-likelihood phylogenies: assessing the performance of
389   PhyML 3.0. *Systematic biology* 2010;59:307–321.

390 45. **Rambaut A, Drummond A**. FigTree: Tree figure drawing tool, version 1.2. 2. *Institute of*
391   *Evolutionary Biology, University of Edinburgh*.

392 46. **Paraskevopoulou S, Käfer S, Zirkel F, Donath A, Petersen M, *et al.*** Viromics of extant insect
393   orders unveil the evolution of the flavi-like superfamily. *Virus Evolution* 2021;7:veab030.

394 47. **Culley AI, Lang AS, Suttle CA**. High diversity of unknown picorna-like viruses in the sea.
395   *Nature* 2003;424:1054–1057.

396 48. **Zhang Y-Y, Chen Y, Wei X, Cui J**. Viromes in marine ecosystems reveal remarkable
397   invertebrate RNA virus diversity. *Sci China Life Sci*. Epub ahead of print 17 June 2021. DOI:
398   10.1007/s11427-020-1936-2.

399 49. **Lang AS, Vlok M, Culley AI, Suttle CA, Takao Y, *et al.*** ICTV Virus Taxonomy Profile:
400   Marnaviridae 2021. *Journal of General Virology* 2021;102:001633.

401 50. **Brussaard CPD, Noordeloos AAM, Sandaa R-A, Heldal M, Bratbak G**. Discovery of a dsRNA
402   virus infecting the marine photosynthetic protist Micromonas pusilla. *Virology*
403   2004;319:280–291.

404 51. **Day JM**. The diversity of the orthoreoviruses: molecular taxonomy and phylogentic divides.
405   *Infection, Genetics and Evolution* 2009;9:390–400.

406 52. **Valles SM, Chen Y, Firth AE, Guérin DMA, Hashimoto Y, *et al.*** ICTV virus taxonomy profile:
407   Iflaviridae. *The Journal of General Virology* 2017;98:527.

408 53. **Guo L, Lu X, Liu X, Li P, Wu J, et al.** Metatranscriptomic Analysis Reveals the Virome and
409   Viral Genomic Evolution of Medically Important Mites. *Journal of Virology*;95:e01686-20.

410 54. **Steil BP, Kempf BJ, Barton DJ**. Poly(A) at the 3️ End of Positive-Strand RNA and VPg-Linked
411   Poly(U) at the 5️ End of Negative-Strand RNA Are Reciprocal Templates during Replication
412   of Poliovirus RNA. *Journal of Virology* 2010;84:2843–2858.

413 55. **Gann ER, Kang Y, Dyhrman ST, Gobler CJ, Wilhelm SW**. Metatranscriptome Library
414   Preparation Influences Analyses of Viral Community Activity During a Brown Tide Bloom.
415   *Frontiers in Microbiology* 2021;12:1126.

416 56. **Canuti M, van der Hoek L**. Virus discovery: are we scientists or genome collectors? *Trends*
417   *in microbiology* 2014;22:229–231.

418 57. **Greninger AL**. A decade of RNA virus metagenomics is (not) enough. *Virus Research*
419   2018;244:218–229.

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434   **Figure 1: Summary of viral contigs discovered from echinoderm transcriptomes** (A) The
435   number of transcriptomes downloaded from NCBI ordered by echinoderm class that were
436   processed for viral discovery. (B) Top bars display the total number of viral contigs discovered
437   separated by echinoderm class and life stage. Bottom bars display total number of transcriptomes
438   separated by echinoderm class and life stage (C) Percentage of viral reads in transcriptomes by
439   life stage. NS = non-significant.

440

441
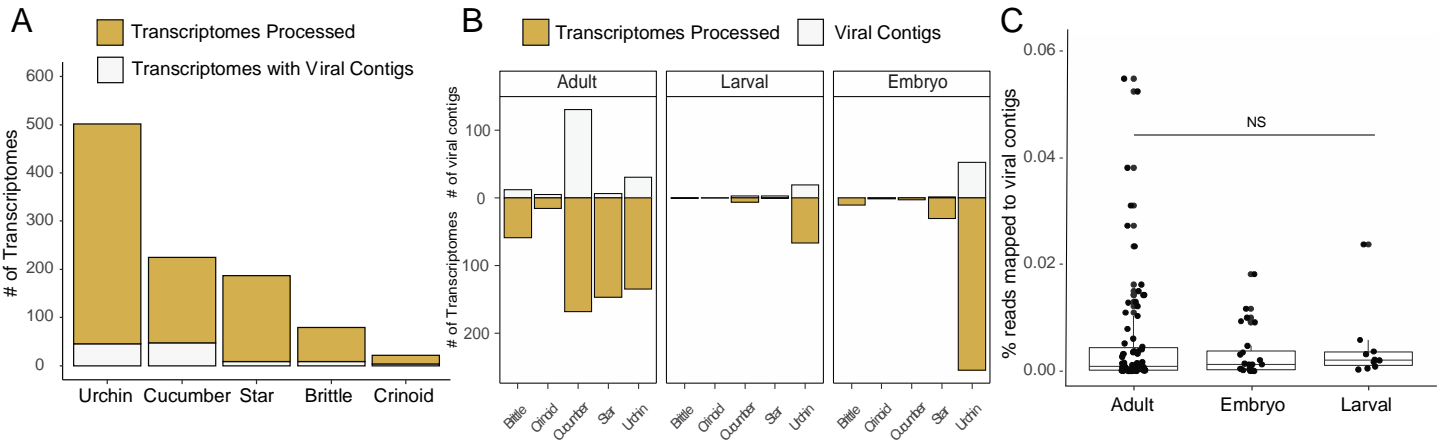
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467

Figure 2: Picornaviruses are the dominant viral order found in echinoderms. (A) Colored circles 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.
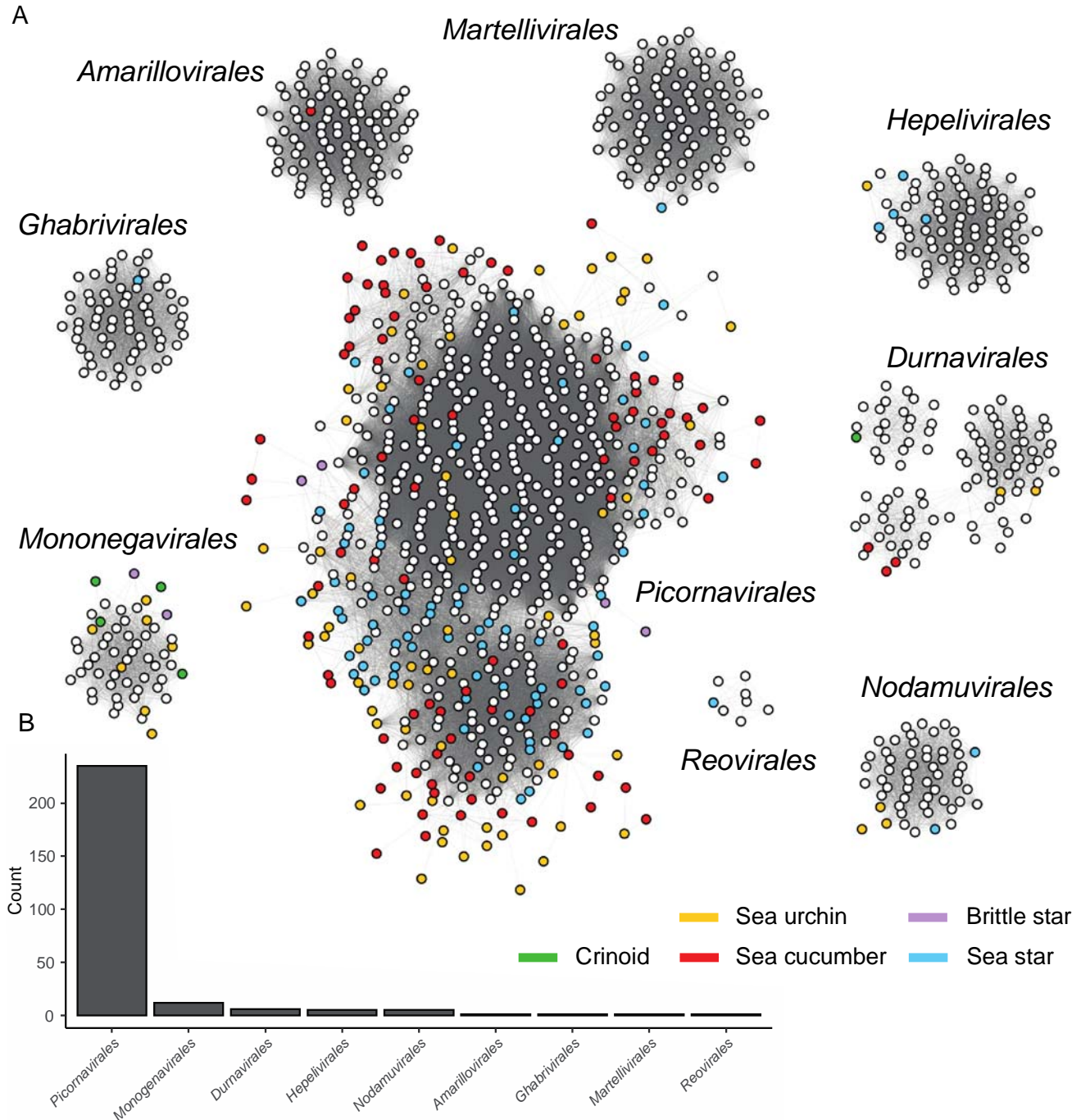
Figure 3: Echinoderm picornaviruses are broadly distributed across the *Picornavirales* phylogeny. Tips are colored by taxonomic family with orange circles representing echinoderm picornaviruses. Genome architectures of complete and near complete genomes recovered from assemblies displayed**.** Genomes are drawn approximate to scale in a 5' to 3' direction. Open reading frames denoted by boxes and colored by general function. Asterisk represents an incomplete open reading frame. Animal icons represent the echinoderm order the viral contig is associated with.
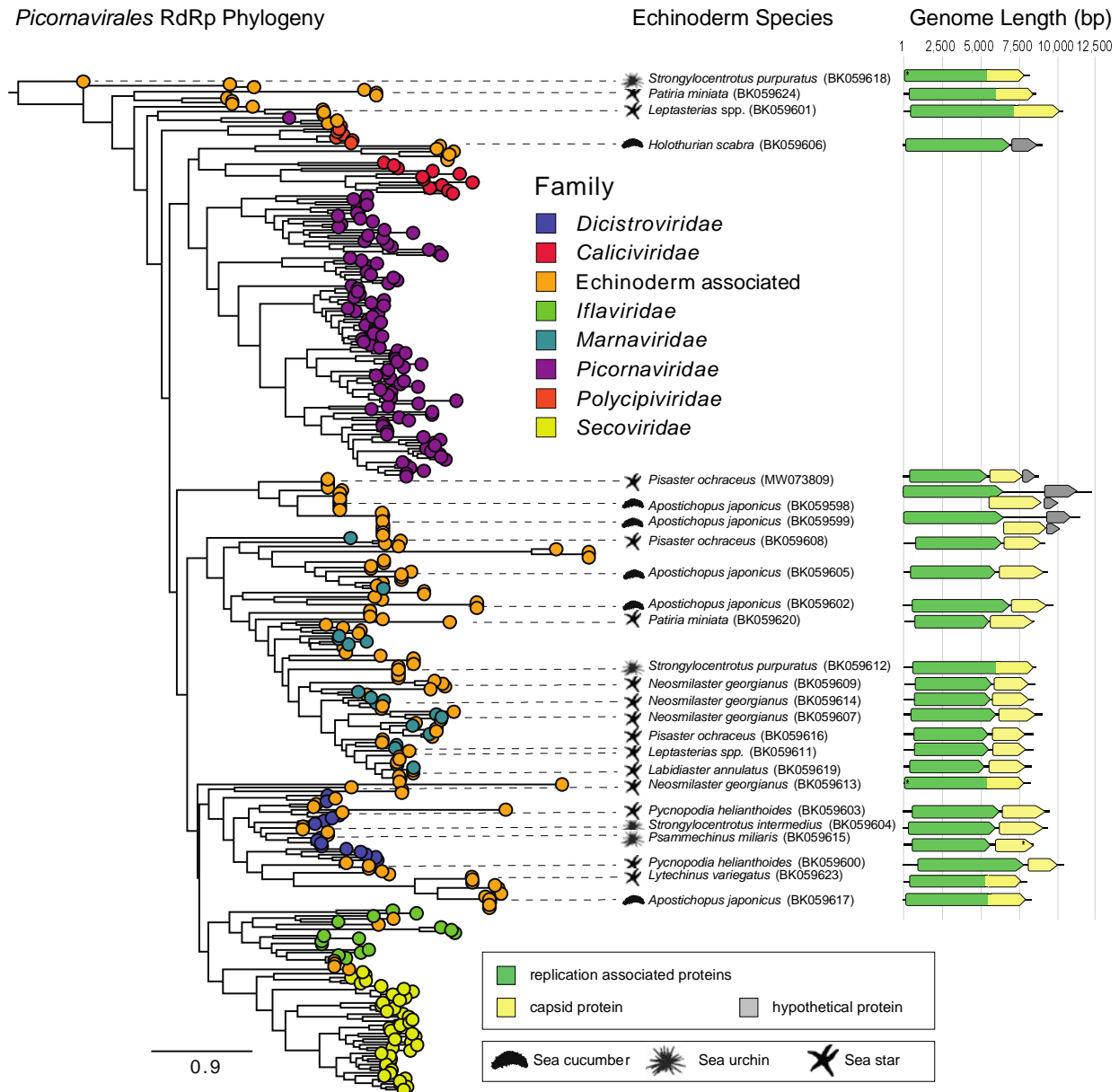
Figure 4: Phylogenetic placement of echinoderm viruses from *Reovirales, Mononegavirales, Amarillovirales,* and *Hepevirales*. Tips are colored by taxonomic family or genus with black circles representing echinoderm viruses. Genome architectures of complete and near complete genomes recovered from assemblies displayed**.** Genomes are drawn approximate to scale in a 5' to 3' direction. Open reading frames denoted by boxes and colored by general function. Asterisk represents an incomplete open reading frame and a question mark denotes potentially incomplete
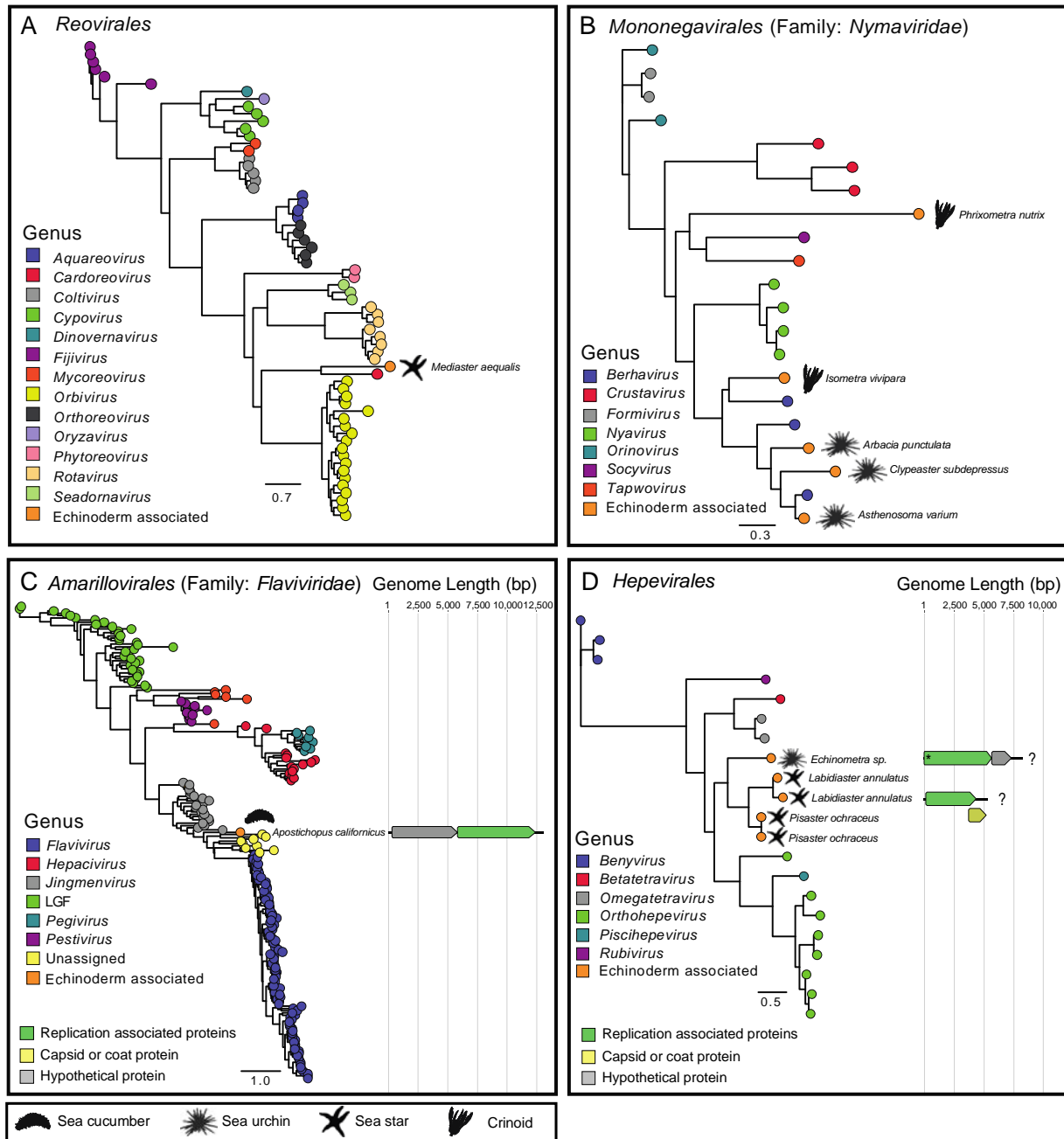
genome. Animal icons represent the echinoderm order the viral contig is associated with.

Figure 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 colored 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.
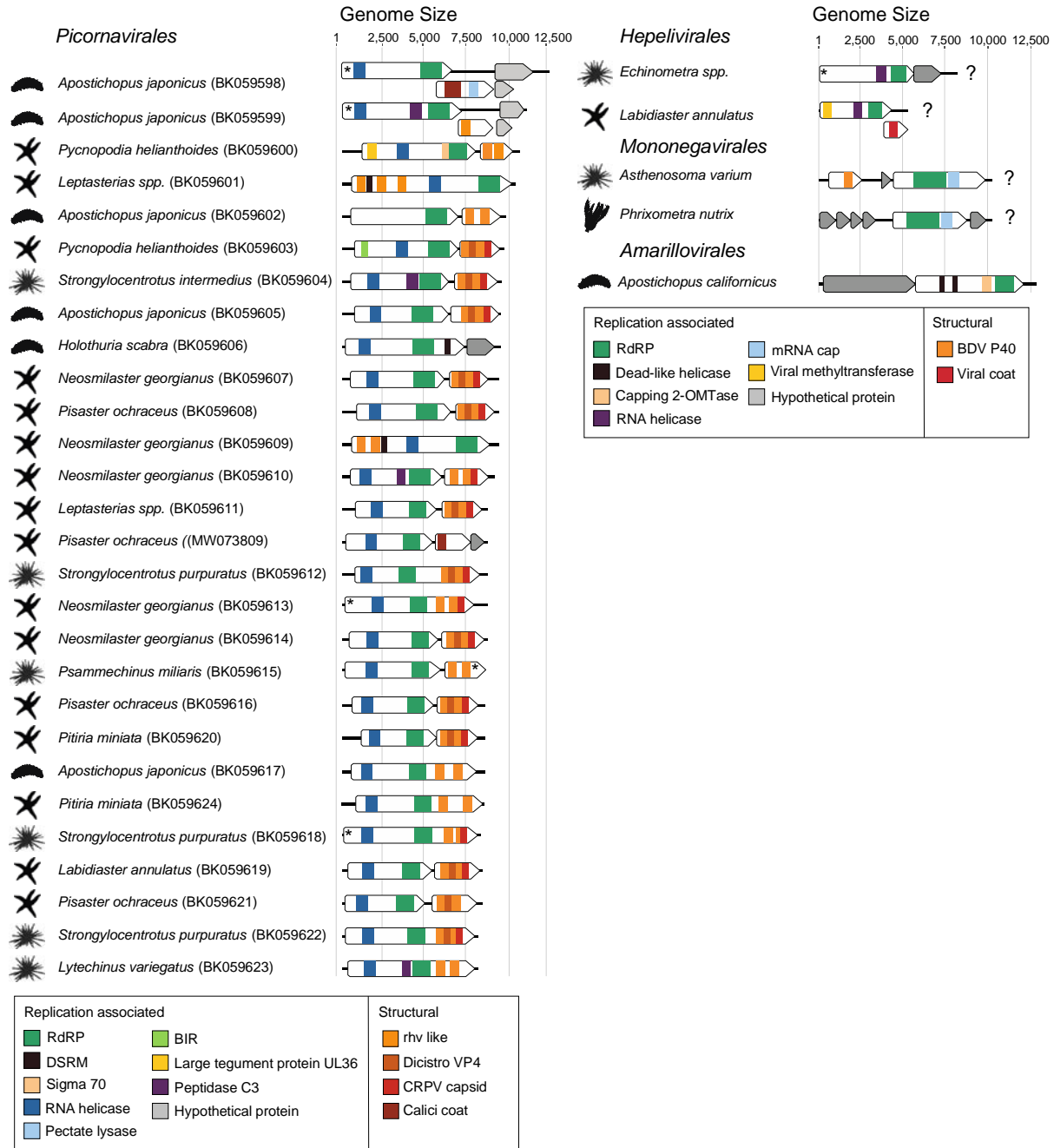
500 **Figure S1: Summary statistics of viral discovery from short read libraries.** (A) The
501 distribution of viral contig lengths between RNA viral metagenomes and transcriptomes. (B)
502 Total number of viral contigs discovered in RNA viral metagenomes and host transcriptomes.
503 (C) Pearson's correlations of total reads in a given library and total viral sequences in the same
504 library.
505
506
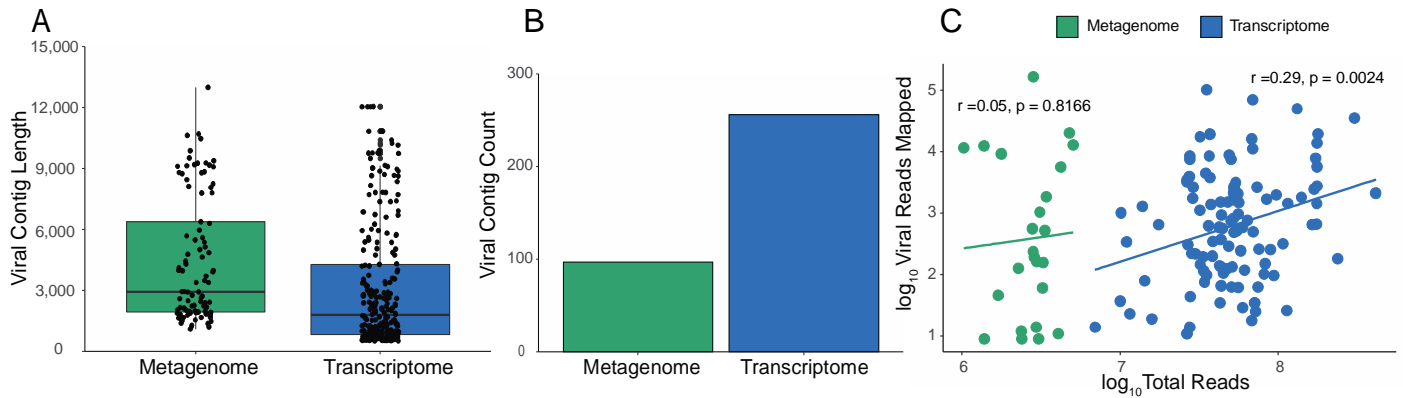
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533