

Widespread occurrence and diverse origins of polintoviruses influence lineage-specific genome dynamics in stony corals

Danae Stephens, Zahra Faghihi, and Mohammad Moniruzzaman^{†,*}

Department of Marine Biology and Ecology, The Rosenstiel School of Marine, Atmospheric and Earth Science, University of Miami, 4600 Rickenbacker Causeway, Miami, FL 33149-1031, USA

[†]<https://orcid.org/0000-0001-9337-3874>

*Corresponding author: E-mail: m.monir@miami.edu

Abstract

Stony corals (Order: *Scleractinia*) are central to vital marine habitats known as coral reefs. Numerous stressors in the Anthropocene are contributing to the ongoing decline in coral reef health and coverage. While viruses are established modulators of marine microbial dynamics, their interactions within the coral holobiont and impact on coral health and physiology remain unclear. To address this key knowledge gap, we investigated diverse stony coral genomes for ‘endogenous’ viruses. Our study uncovered a remarkable number of integrated viral elements recognized as ‘Polintoviruses’ (Class *Polintoviricetes*) in thirty *Scleractinia* genomes; with several species harboring hundreds to thousands of polintoviruses. We reveal massive paralogous expansion of polintoviruses in stony coral genomes, alongside the presence of integrated elements closely related to Polinton-like viruses (PLVs), a group of viruses that exist as free virions. These results suggest multiple integrations of polintoviruses and PLV-relatives, along with paralogous expansions, shaped stony coral genomes. Re-analysis of existing gene expression data reveals all polintovirus structural and non-structural hallmark genes are expressed, providing support for free virion production from polintoviruses. Our results, revealing a significant diversity of polintovirus across the *Scleractinia* order, open a new research avenue into polintovirus and their possible roles in disease, genomic plasticity, and environmental adaptation in this key group of organisms

Keywords: *Scleractinia*; coral-virus interactions; polintons; polintovirus; polinton-like virus; endogenous viruses; integrated viral elements; coral; stony coral.

Introduction

Coral reefs are a striking feature of many marine habitats, glowing with an abundance of life of every size, shape, and color. These reefs are formed by a group of clonal composite symbiotic organisms (Order: *Scleractinia*) with individual colonies composed of genetically identical polyps hosting single celled algae of the *Symbiodiniaceae* family. The symbiont algae provide energy to their coral hosts, which secrete and cement calcium carbonate skeletons into vast three-dimensional structures (Knowlton and Rohwer 2003; LaJeunesse et al. 2018; Matthews et al. 2020). Globally, coral reefs have been estimated to harbor 25 per cent of marine biodiversity (Thurber et al. 2017) and even beyond that scope; through recreation, tourism, and natural resource use, they are estimated to provide around \$2.7 trillion USD worldwide in economic value each year (Souter et al. 2021). However, for as much as we value these reefs and their socioeconomic contributions, coral reef health and coverage are in decline, with coral coverage dropping from 33.3 per cent in 2009 to 28.8 per cent in 2018 (Souter et al. 2021), with that declining trend only accelerating in recent years.

There are multiple factors impacting this decline, which include warming oceanic temperatures (inducing longer lasting, higher intensity, and more frequent long-term mass bleaching events and their associated mortality) (Knowlton 2001; Hughes et al. 2003; Pandolfi et al. 2003; Hoegh-Guldberg et al. 2007; Souter et al. 2021) decreased pH (and the associated lowered calcification rate slowing colony growth) (Hoegh-Guldberg et al. 2007; Souter et al. 2021), and the rapid spread and transport of coral diseases across colonies and between reef sites (Pandolfi et al. 2003; Richardson 2004; Souter et al. 2021). These factors have different levels of impact on coral reefs across the world, accounting for the main stressors in any given local environment. For example, in the Caribbean specifically, it is disease that plays the largest role in coral mortality and reef decline (Gardner et al. 2003). Of the great number of disease outbreaks occurring across corals over the years (Richardson 1998; Pollock et al. 2011; Mera and Bourne 2017), the most well-known and devastating outbreaks are those that occurred in the Caribbean (Green and Bruckner 2000; Gardner et al. 2003; Cróquer and Weil 2009; Muller and van Woesik 2012; Woesik and Randall 2017). These various diseases have originated

due to a number of factors, with many diseases having multiple compounding causal agents (Richardson 1998; Mera and Bourne 2017), viruses being one of them.

Virus–host interactions in coral have a number of effects, from the more traditional role of viruses as agents of infection and disease (Duerkop and Hooper 2013; Volkman and Stetson 2014; Wiersinga et al. 2014; Grasis et al. 2017), to alterations of the host holobiont and its associated communities impacting host health and resilience (Wiersinga et al. 2014; Grasis et al. 2017). In the context of coral holobionts, the role of viruses could be both beneficial and harmful. These effects potentially encompass pathogenic interactions, as well as important top-down control mechanisms that constrain the microbiome community dynamics (Aswad and Katzourakis 2012; Blanc, Gallot-Lavallée, and Maumus 2015), in addition to interactions that potentially result in increased immunity and protection from other pathogens (Barton, White, and Virgin 2009; Blanc, Gallot-Lavallée, and Maumus 2015; Grasis et al. 2017; Mougari et al. 2020; Hackl et al. 2021).

So far, multiple approaches have been used to examine viral interactions with coral, from transmission electron microscopy in early reports examining the presence of virus-like particles within coral tissue (Wilson et al. 2005; DAVY and Patten 2007), to metagenomic approaches examining both coral and symbiont strains in diverse locations and under diverse environmental conditions (Dinsdale et al. 2008; Webster et al. 2016; Thurber et al. 2017; Weynberg et al. 2017). Despite this, there has been little work done to determine the function and identity of most members of the holobiont virus communities within coral (Weynberg et al. 2017), which remains an important task given the key contribution of viruses in diverse ecosystems.

One underexplored aspect of virus–host interactions with potential implications for coral health and holobiont stability is the level of viral endogenization within coral host genomes. Endogenization refers to the integration of near-complete or partial viral genomes into the host genome. Endogenous viruses exist in varying levels within the genomes of numerous eukaryotes (Feschotte and Pritham 2007; Holmes 2011; Liu et al. 2011), and the process of endogenization has been ongoing for millions of years throughout the course of eukaryotic evolution (Gilbert, Feschotte, and Emerman 2010; Katzourakis, Gifford, and Malik 2010; Worobey et al. 2010; Aswad and Katzourakis 2012). These elements can be incredibly diverse, with instances of hundreds to thousands of such elements inhabiting a single genome (Hackl et al. 2021; Bellas et al. 2023), and large-scale endogenization of near-complete genomes of ‘giant viruses’, comprising a high percentage of the total genome content in many eukaryotic species (Feschotte and Pritham 2007; Holmes 2011; Liu et al. 2011; Moniruzzaman et al. 2020).

Endogenous viruses reveal the history of past interactions of hosts with viruses, as the endogenization process leaves traces of past infections in the genomes of their hosts (Patel, Emerman, and Malik 2011; Aswad and Katzourakis 2012; Blanc, Gallot-Lavallée, and Maumus 2015; Moniruzzaman et al. 2020). Thus, endogenous viral elements (EVEs) can provide an opportunity to reveal the phylogenetic identity of extant viruses that infect a particular host lineage. A notable, relevant study in this regard is the identification of RNA viruses infecting the Symbiodiniaceae members of coral holobionts based on RNA virus EVE signatures in *Symbiodinium* genomes (Howe-Kerr et al. 2022).

Motivated by the lack of information on the interaction of viruses with corals, we performed a comprehensive investigation of the stony coral genomes for endogenous virus signatures with a goal to shed light on the roles of such elements in coral

genome evolution along with identifying the viral lineages that likely infect corals in nature. Intriguingly, our study led to the identification of ‘Polintoviruses’ residing within the genomes of stony, reef building coral (Order: *Scleractinia*) from eleven different genera. Polintoviruses, also known as Maverick/Polinton elements, were first reported in 2005 (Rinke and Jurka et al. 2006). Originally defined as transposable elements (selfish genetic elements that can proliferate within the genome of the host) (Calos and Miller 1980; Finnegan 1992; Yutin, Raout, and Koonin 2013; Bourque et al. 2018), these are large elements approximately 10 kb to 50 kb in size (Haapa-Paananen, Wahlberg, and Savilahti 2014; Bellas and Sommaruga 2021), representing some of the largest integrating DNA elements known. Later discovery of conserved structural proteins (major and minor capsid proteins; MCP and mCP, respectively) in these elements prompted re-classification of these elements as viruses (Krupovic, Bamford, and Koonin 2014), as capsid proteins are considered to be a key and defining characteristic of viruses 51 and it was determined that all known polinton sequences harbored these capsid proteins. This classification of polintoviruses as bona fide viruses is officially recognized by the International Committee on Taxonomy of Viruses (ICTV). Specifically, these viruses are included in the family Adintoviridae, which belongs to the order Orthopolintovirales within the class Polintoviricetes52.

Further study of the MCP of polintoviruses determined it to be closely related to the MCPs of virophage (Yutin et al. 2015), a class of small dsDNA viruses that need Nucleocytoplasmic large DNA viruses (NCLDVs) for replication (Stough et al. 2019). This led to the hypothesis that virophages originated from Polintoviruses (Krupovic, Bamford, and Koonin 2014). This enhanced understanding led to the discovery of Polinton-like viruses (PLVs) in 2015 (Yutin et al. 2015). PLVs as well as a few close relatives of PLVs that harbor polintovirus-like characteristics were found to be highly abundant in aquatic environments (Bellas and Sommaruga 2021).

In this study, we report the widespread prevalence of polintoviruses and relatives of PLVs within diverse stony coral genomes, and analyze their phylogenetic histories, gene expression dynamics, and functional potential. Our study opens the door for investigations into the origins of these elements and their impact on coral physiology, environmental adaptation, and genome evolution.

Results and discussion

Prevalence of polintoviruses in stony coral genomes

Using the bioinformatic pipeline we established for screening viral regions originating from polintoviruses (see ‘Materials and methods’ section), PLVs, and virophages, we identified numerous endogenous viral regions (EVRs; Fig. 1B) within each of the thirty analyzed *Scleractinia* genomes (genome accession numbers in Supplementary Table S1). Our preliminary screening utilized a strict cut-off criteria dictating that each region needed to contain a major capsid protein (MCP), minor capsid protein (mCP), and ATPase gene—genes that are commonly present and conserved across the members of class Polintoviricetes, virophages, and PLVs (Koonin and Krupovic 2017; Bellas and Sommaruga 2021). These EVRs varied in length widely within our specified size cut-off (10–50 kb; Supplementary Fig.S1) consistent with what has been previously found in several studies (Starrett et al. 2021; Bellas et al. 2023). The overall mean and median length for validated EVRs were found to be 23,555 base pairs (bp) and 21,932

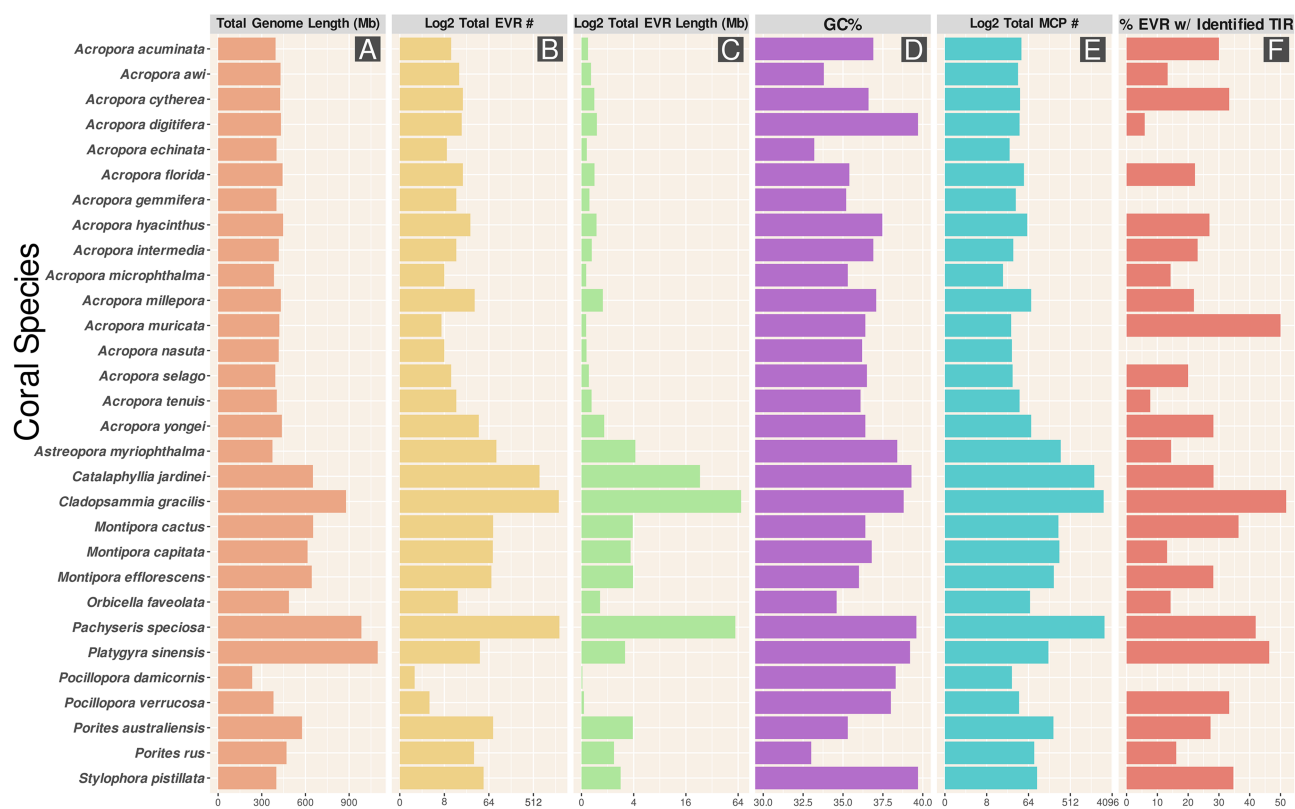


Figure 1. Bar plots of coral species with regard to (A) total genome length, measured in million base pairs (Mb), (B) total number of endogenized viral regions (EVR), represented on log₂ scale, (C) combined length of all EVRs within a genome measured in megabase pairs (mb) and represented on a log₂ scale, (D) overall percentage of GC content of EVRs in each coral genome, (E) total number of identified MCP genes per coral genome, represented on log₂ scale, (F) proportion of EVRs that harbor TIRs.

bp, respectively, with an overall right-skewed distribution of EVR lengths (Supplementary Fig. S2).

When we utilized a more relaxed quality control criteria (set to include any three hallmark genes, regardless of type), we found that on average 26.78 per cent more putative EVRs passed cut-off requirements (Paired *t*-test, *P*-value = 0.0212). This indicates that our results represent a conservative estimate of polintoviruses in stony coral genomes. Due to the understudied nature of polintoviruses, it is likely that substantial sequence diversity exists in some of their core genes that render them difficult to be accurately identified with existing information on their hallmark genes, such as the highly divergent MCP homologs identified by Bellas et al. (2023). Likewise, the sequencing methods utilized in the genome we investigated might have resulted in fragmented and incomplete viral regions, further limiting our ability to comprehensively identify and isolate these EVR. While we chose to use stringent quality controls to ensure the validity of our initial exploratory results, we note that many of the EVRs excluded may in fact be *bona fide* viral candidates.

In addition to identifying the viral regions, we also searched specifically for MCP homologs (within or outside the viral regions) across the genomes, and we found consistently higher numbers of MCPs in most genomes than validated EVRs (Fig. 1B,E). We found that number of MCPs in the stony coral genomes were on average 34.7 per cent more than the number of EVRs, with the largest discrepancy between these counts observed in *Cladopsammia gracilis* (*C. gracilis*) (62.4 per cent more MCP hits than EVR counts) and the lowest *Pocillopora damicornis* (*P. damicornis*) genome (3.6 per cent). This supports the notion that the ‘true’ number of polintovirus

EVR in stony coral species is likely higher than what we have found. This observation is also supported by Bellas et al. (2023), who found that the integrated MCPs are usually part of endogenous polintoviruses, which might not always be detectable due to low quality genome assemblies from short reads.

This discrepancy between the number of EVR and MCP candidates recovered may be due, in part, to the fact that the coral genomes analyzed in our study used varying sequencing and assembly approaches. It has been demonstrated that short read sequencing techniques, such as Illumina (Loch et al. 2008) (used for all but three of the genomes studied, which were assembled using PacBio data), are limited in their ability to capture the entirety of endogenous viruses. This is due to the process resulting in fragmented assemblies of eukaryotic genomes, which are often full of repetitive elements, along with miss-assemblies, thus introducing frameshifts or other errors that make identification of such regions difficult. In addition, fragmented assemblies often lead to small contigs which might contain one or two viral hallmark genes but will not represent the entire EVR. This possibility is supported by a recent study by Bellas et al. (2023), who have reported recovery of complete endogenized PLVs in draft genomes after reassembly and error-correction of several protist genomes using long-read sequencing technology.

In further support of this, an additional study by Noel et al. (2023) determined that long read sequences were essential for developing and obtaining an accurate understanding of gene content, as their work in assembling coral genomes from Tara Pacific (Lombard et al. 2023) data demonstrated that a great deal of ‘repeat’ and duplicated genetic elements were lost using short

read data. Taken together with the findings of [Bellas et al. \(2023\)](#), it is highly likely that contig fragmentation, gene loss, or divergent copies of core genes are responsible for these ‘missing’ core genes in our identified EVR. Nevertheless, we found a strong positive correlation between the number of EVR ([Fig. 1B](#)) and the number of MCP hits ([Fig. 1E](#)) across genomes, (Spearman’s $\rho=0.9467648$), indicating that while the number of EVR we identified might be conservative, it is nevertheless a good indicator of the genus or species-specific variations in polintovirus abundances across the stony coral genomes.

We also sought to identify the presence of Terminal Inverted Repeats (TIRs) within our EVRs. TIRs are hallmarks of inserted elements ([Munoz-Lopez and Garcia-Perez 2010](#)), and previous studies have found that in protist genomes, as much as 50 per cent of the polintoviruses located on long contigs can harbor TIRs at their ends ([Bellas et al. 2023](#)). We found that many of our identified EVR candidates harbor TIRs, with the highest percentage of validated EVRs containing TIRs was found within *C. gracilis*; wherein 869 of the 1,676 validated EVRs possessed TIRs ([Fig. 1F](#); [Supplementary Fig. S3](#)). On average 22.8 per cent of the identified EVR across the coral genomes harbored TIRs. It is likely that we were unable to detect additional TIRs-containing EVRs given the genome assembly quality—with many of these regions residing on small contigs.

Overall, some of the stony coral genomes harbored a remarkably high number of EVRs and MCPs. Of note are the genomes of *C. gracilis* and *Pachyseris speciosa* (*P. speciosa*), with the former harboring 1,676 EVRs and 2,686 MCPs, and the latter harboring 1,724 EVRs and 2,802 MCPs. All other genomes examined contained less than 100 EVRs and 320 MCP hits; except for *Catalaphyllia jardinei* (*C. jardinei*), which harbored 680 EVRs and 1,685 MCP hits. The genome with the least number of hits was *P. damicornis*, harboring only 1 EVR that we were able to identify within our specified criteria ([Fig. 1B](#); [Supplementary Fig. S4](#)). These results reveal the remarkable variations in prevalence of polintoviruses across different members of the *Scleractinia* order.

The smallest of the genomes we analyzed was *P. damicornis* (234.35 Mbp), and the largest was *Platygyra sinensis* (*P. sinensis*) (1,096.36 Mbp) ([Fig. 1A](#); [Supplementary Fig. S6](#)). Despite the limitations of the genome sequencing methods utilized, the completeness for the *Scleractinia* genomes was high, the average BUSCO completeness score 89.3, 63 per cent of the genomes containing a BUSCO score over 90, 83 per cent with a score over 85 per cent, and the *Porites rus* genome having the lowest completeness score at 70.2 per cent. The relatively high completeness of most of the genomes ensures that we were able to detect the majority of the integrated viruses in these genomes. In support of this, we identified a positive correlation between length of coral genome and number of EVR (Spearman’s $\rho=0.73$) ([Supplementary Fig. S7](#)). We likewise found a positive correlation between the length of the coral genome and the number of MCP hits ([Fig. 1E](#); [Supplementary Fig. S8](#); Spearman’s $\rho=0.69$). These, combined with the high positive correlation between number of EVR and number of MCP hits (Spearman’s $\rho=0.95$) together provide their own level of confidence on our quality control. As this study was primarily focused on assessing the diversity of EVEs across different species of stony corals, we did not attempt to further normalize the results to genome completeness.

Positive correlation between the genome length and number of EVRs suggests that at least in the limited subset of genomes we analyzed, longer genomes tend to harbor larger numbers of EVRs. Interestingly, we found that *P. sinensis*, despite having the longest genome (at 1,096 Mbp), contained only 41 EVR and 172

MCP hits, whereas *P. speciosa* at the second longest (984 Mbp) held the second highest number of EVRs ([Fig. 1A,B,E](#)). This observation suggests that factors other than genome sizes are also responsible for the lineage-specific variations in the EVR abundances we observed.

Due to these vast numbers of EVRs, in several assemblies they appear to contribute a sizable portion of the genomes ([Fig. 1C](#))—e.g. ~ 7.8 per cent of the *C. gracilis* genome is composed of these EVRs ([Supplementary Fig. S5](#)). Indeed, if we consider the possibility that most of the MCP hits in this genome are part of endogenous polintoviruses, then the true contribution of polintoviruses to the genome content of *C. gracilis* will be even higher. Although relatively rare, similar massive proliferation of polintoviruses has been reported in protist *Trichomonas vaginalis*, where approximately 27–54 per cent of the genome length was contributed by these elements [Bellas et al. \(2023\)](#). The same study also reported that ~ 10 per cent of the genome of *Paulinella micropora* consists of PLVs. As far as we are aware, *P. speciosa* has the largest number of polintovirus regions ever reported in metazoans.

Phylogeny of the stony coral polintovirus genomes

For phylogenetic analysis, we first clustered the MCPs from the identified EVRs at a 90 per cent amino acid sequence similarity to reduce redundancy in the dataset (see ‘Materials and methods’ section). In doing so, we determined that the number of MCP clusters is quite high in most genomes, indicating high within-genome diversity of polintoviruses in stony corals. A similar level of diversity of PLVs was reported in several protist genomes, indicating the possibility of multiple independent integration events shaping the diversity within these genomes ([Bellas et al. 2023](#)). We then performed a maximum likelihood phylogenetic assessment on the representative sequences from each of these clusters along with reference *Adintoviridae* members, PLV, and viroplage MCPs (see ‘Materials and methods’ section) to clarify the evolutionary history of the EVRs in stony corals ([Fig. 2](#); [Supplementary Fig. S9](#)).

We found that the vast majority of the stony coral EVRs clustered separately from the environmental PLVs and viroplages ([Fig. 2](#)), supporting previous observations ([Koonin and Krupovic 2017](#); [Barreat and Katzourakis 2023](#)) regarding the evolutionary history of these elements. Most of the environmental viroplages and PLVs formed their own distinct clusters on the phylogeny, although a few of the PLVs did cluster with the viroplages. We used a recently proposed homology-based approach to identify the viroplages in our datasets ([Roux et al. 2023](#)) (see ‘Materials and methods’ section), and it is possible that a few false-positives emerged in this classification approach. In contrast, the majority of the stony coral polintoviruses formed separate clusters along with the members of the previously established *Adintoviridae* family. *Adintoviruses* were reported within diverse animal genomes, which have since been identified to be essentially polintoviruses, and thus were later classified within the class *Polintoviricetes* ([Walker et al. 2021](#)).

Polintoviruses within the stony coral genomes formed two large monophyletic clusters that shared a common ancestor ([Fig. 2](#)). Each of these clusters harbored polintoviruses from most of the coral genera, suggesting presence of phylogenetically distinct polintoviruses within the same lineage (genus or species) of stony coral ([Fig. 2](#)). This is consistent with multiple origin of these elements throughout the evolutionary history of the *Scleractinia* order. Phylogenetic analysis showed distinct patterns of paralogous expansion of stony coral polintovirus genomes. Specifically,

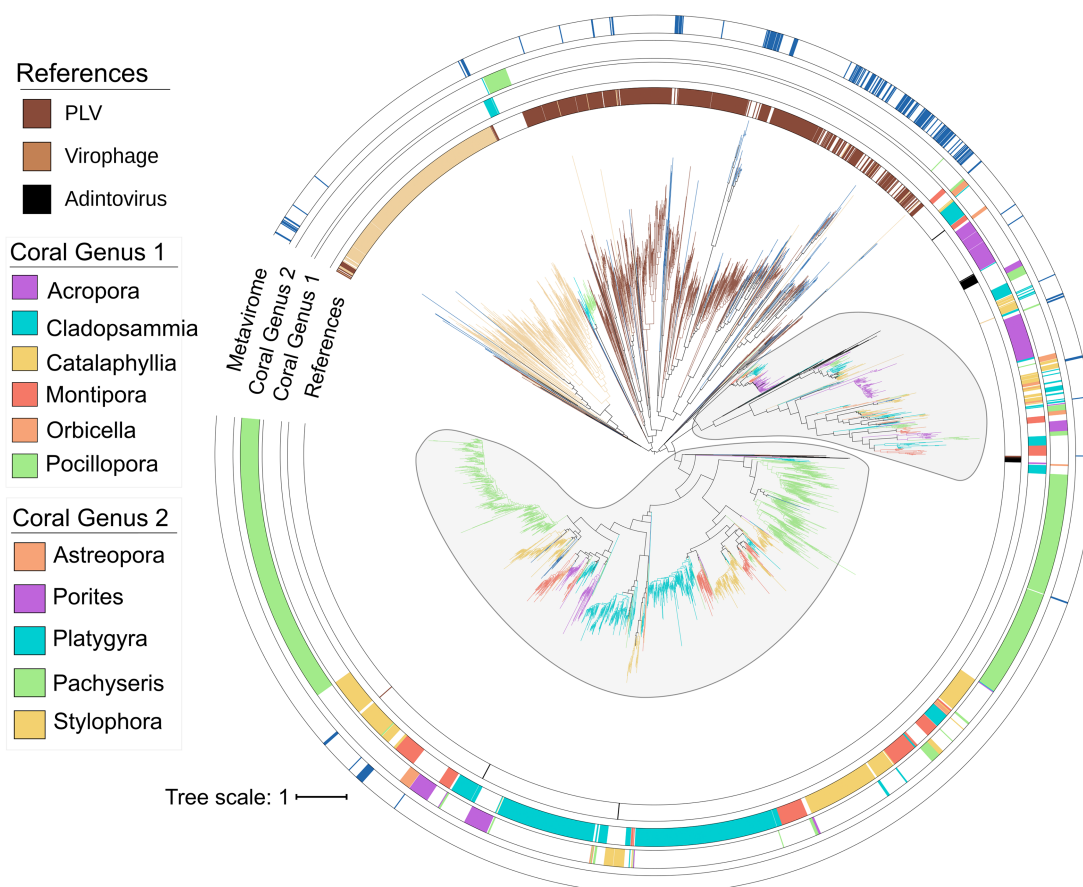


Figure 2. Maximum likelihood phylogeny representing the diversity and evolutionary history of the reference PLVs, PLVs in the coral virome, and stony coral EVRs. The tree was reconstructed using MCPs from reference viral genomes that are established members of the Adintoviridae family, polinton-like viruses (PLVs), and virophages, alongside MCPs from EVRs found in the genomes of stony corals. In addition, PLV-specific MCPs from a previous coral virome study (Weynberg et al. 2017) are also included in this phylogeny. The two large monophyletic clusters that harbor the polintoviruses from diverse stony coral genomes are shaded in grey. The phylogenetic tree is annotated across four concentric rings for clarity. Ring 1 (innermost): MCPs of reference virus sequences. Rings 2 and 3: MCPs from EVRs identified in stony corals that are annotated by different coral genera. Ring 4 (outermost): MCPs recovered from stony coral virome.

a large clade within the tree where the majority of the EVRs are clustered (except *Acropora*) showed large expansion of these elements. The largest degree of such expansion was found to be in the species *C. gracilis*, *C. jardenei*, and *P. speciosa*, consistent with the fact these three species harbored the largest number of these elements overall (Fig. 1B). A separate cluster of stony coral polintoviruses that also harbored representatives from most genus did not show as much paralogous expansion compared to the cluster previously discussed. All the EVRs from different *Acropora* species are also clustered within this group. We observed a similar phenomenon of genus-specific clustering of EVRs for other coral genomes for which we were able to analyze multiple species; specifically, *Montipora* ($n=3$), *Porites* ($n=2$), and *Pocillopora* ($n=2$). While EVRs from the same genus generally cluster together, we also found several instances where EVRs from multiple genus clustered within close proximity to each other, suggesting exchange of these elements across different genera. Another possibility is that ancestral infection followed by genome integration and paralogous expansion of some of these polintovirus elements occurred before eventual speciation within some genera.

To date, a comprehensive assessment of the evolutionary origin of polintoviruses remains to be performed, despite their identification in a number of vertebrate and invertebrate lineages (Feschotte and Pritham 2007; Holmes 2011; Liu et al. 2011). Our observation

of distinct patterns of expansion of the stony coral polintoviruses shaped by their phylogenetic history suggests that while some of these elements went through within-genome amplification upon integration, others were not subject to such expansion to a similar degree. Paralogous expansion of endogenized polintoviruses constrained by the phylogenetic history might provide some type of fitness benefit to the host—a possibility that needs further scrutiny. Future studies focusing on the molecular underpinnings of the proliferation dynamics of these elements could address questions regarding the genomic and evolutionary factors shaping the dynamic expansion landscape of polintoviruses in stony corals.

Even though the vast majority of the stony coral EVRs clustered distinctly from the environmental PLVs, we observed clustering of several EVRs with the established phylogenetic group of environmental PLVs (Fig. 2). Specifically, a number of the elements from the *Cadopsammia*, *Pachyseris*, and *Platygyra* genera formed a deeply branching clade situated between the majority of the reference PLVs and the majority of the reference virophages. A handful of PLVs clustered within this group, including Yellowstone Lake virophage 6. This virophage was previously reported to be a part of the PGVV group of virophages53.

As discussed previously, *Cladopsammia* and *Platygyra* also harbor a large number of EVRs in general, along with high paralogous

expansion. The environmental PLV sequences in our tree originated from free virus particles (Bellas et al. 2023); thus the clustering of some of the stony coral EVRs within this clade supports the possibility that along with polintoviruses, viruses having close evolutionary ties with PLVs potentially infect and integrate within stony coral genomes.

Results from the phylogeny paint a highly dynamic genomic landscape of the stony coral polintoviruses with multiple distinct origins and large-scale lineage-specific expansion likely constrained by their evolutionary histories. The varying degree of expansion of these elements in different stony coral genomes and possible horizontal transfer between genus makes them a key player in the genome evolution of stony corals. Transfer of transposable elements across closely related lineages is not uncommon (Zhuo, Feschotte, and Belshaw 2015). Consistent with our observations, an analysis from a very recent study by Jeong et al. (2023) found the horizontal transfer of polintoviruses across different nematode strains within the same species, and even inter-phylum transfer of these elements. Presence of both phylogenetically distinct polintoviruses and close relatives of PLVs across different stony coral species is consistent with ancestral integration of diverse polintoviruses and PLV-like viruses within host genomes, followed by vertical transfer and lineage-specific proliferation. Further exploration of additional coral genomes as well as a more comprehensive phylogeny will better clarify the evolutionary history of endogenous coral viruses.

Prevalence of PLVs in the virome of stony corals

The abundant and widespread presence of polintoviruses and close relatives of PLVs we uncovered within *Scleractinia* genomes warranted further investigation on whether free viruses, closely associated with polintoviruses or PLVs, could be part of the coral virome. To accomplish this, we leveraged the raw virome sequence data from a study conducted by Weynberg et al. in 2017 (Weynberg et al. 2017). The authors of this study performed a CsCl centrifugation method on several environmental stony coral species to isolate virus-like particles with minimal host contamination to establish the overall viral community. CsCl centrifugation is a widespread technique used in environmental virome studies to isolate virus-like particles (Thurber et al. 2009). This circumvented the methodological limitation of distinguishing virus particles from the coral tissue samples during isolation and sequencing of these components.

Searching this dataset using the MCP Hidden Markov models (HMMs) we constructed from National Center for Biotechnology Information (NCBI) data and Bellas and Sommaruga (2021) (see 'Materials and methods' section) revealed a total of 385 hits to MCP proteins in the virome of several stony coral species. After incorporating these MCPs (following de-replication at >95 per cent amino acid identity, resulting in 322 clusters) within our existing MCP phylogenetic framework, we found that vast majority of these sequences cluster within the environmental PLV clade in our tree, while a few of them were closely affiliated with the virophages (Fig. 2). Interestingly, some of these viral sequences are also clustered within the large clades of stony coral polintoviruses; suggesting that free viruses of high similarity to the polintoviruses could be present in coral viromes. While the possibility of some host contamination cannot be completely ruled out, we do note that the estimated host contamination in these libraries is quite minimal at ≤ 1.5 per cent (Weynberg et al. 2017).

There was not much overlap between the species we analyzed for the EVRs and those sampled for virome by Weynberg et al.

(2017), with only three species (*P. damicornis*, *A. tenuis*, and *P. verucosa*) being common between these two datasets. Despite this limitation, results from the virome revealed PLVs to be an integral component of the stony coral virome, with the potential to infect eukaryotic members of the holobiont, including the coral animal. A targeted approach will be needed in future studies to definitively identify the potential hosts, as well as understand their impact on reef microbial dynamics and impacts on the health and resilience of stony corals themselves.

Functional profile of the stony coral polintoviruses

To understand the functional diversity encoded within the polintovirus EVRs, we clustered the proteins based on sequence similarity and annotated representative proteins from each cluster. We included all the EVRs that were identified using both stringent and relaxed criteria (see 'Materials and methods' section) to explore the full range of functional potential in these regions (Fig. 3A). Apart from the six known core genes that were highly prevalent in these regions, we found several proteins with unknown function to be present in high frequency (>40 per cent of the regions) in stony coral polintoviruses; including a protein homologous to poxvirus protein AMEV120, which contains a Pfam domain (DUF5679) of unknown function. Interestingly, in our analysis, another group of proteins emerged that are present on average in >10 per cent EVRs, which included homologs of putative antirepressor protein YoqD, peptidoglycan binding domain, viral structural proteins, and several proteins with unknown function.

Also of interest, one of these proteins was homologous to the VLTf3 transcription factor, a gene known to be conserved in nucleocytoplasmic large DNA viruses (also known as giant viruses (Moniruzzaman et al. 2020) that infect diverse eukaryotes. The presence of this gene in a subgroup of polintoviruses has been reported before, with suggestions that exchange of this gene occurred between giant viruses and polintoviruses (Krupovic and Koonin 2016). The presence of homologs to antirepressors and additional viral structural proteins further supports the fact that polintoviruses are *bona fide* viruses. However, much is yet to be discovered regarding the role of additional genes in dictating the life-strategies and evolutionary trajectories of polintoviruses. Finally, we found numerous proteins with low frequency (present in >3 per cent of the elements) across the stony coral polintoviruses. Some of these are homologous to antirepressor proteins, helicases, virus structural proteins, exo- and endonucleases, and collagen-like proteins (Supplementary Fig. S10).

The vast majority of the elements that we discovered harbored PolB and retroviral integrases as the genes responsible for DNA replication and integration, respectively. Previous studies have suggested that PLVs, unlike polintons, predominantly harbor DNA helicase and tyrosine recombinases instead for these functions. In our analysis, we found that a handful of the elements (52 out of 4,862) harbor tyrosine recombinases. However, there is no clade-specific pattern as these elements were scattered across the tree. Only one of the coral EVEs that clustered with reference PLVs harbors tyrosine recombinase. Similar distribution patterns were found for different classes of DNA helicases in some of the coral EVEs. Further, in-depth studies will be required for a better assessment of the gene content and their implications in coral EVE replication and possibility of horizontal gene transfer across diverse members of polinton and PLV groups.

Our analysis revealed highly diverse functional profiles encoded across the stony coral polintoviruses, and the patchy

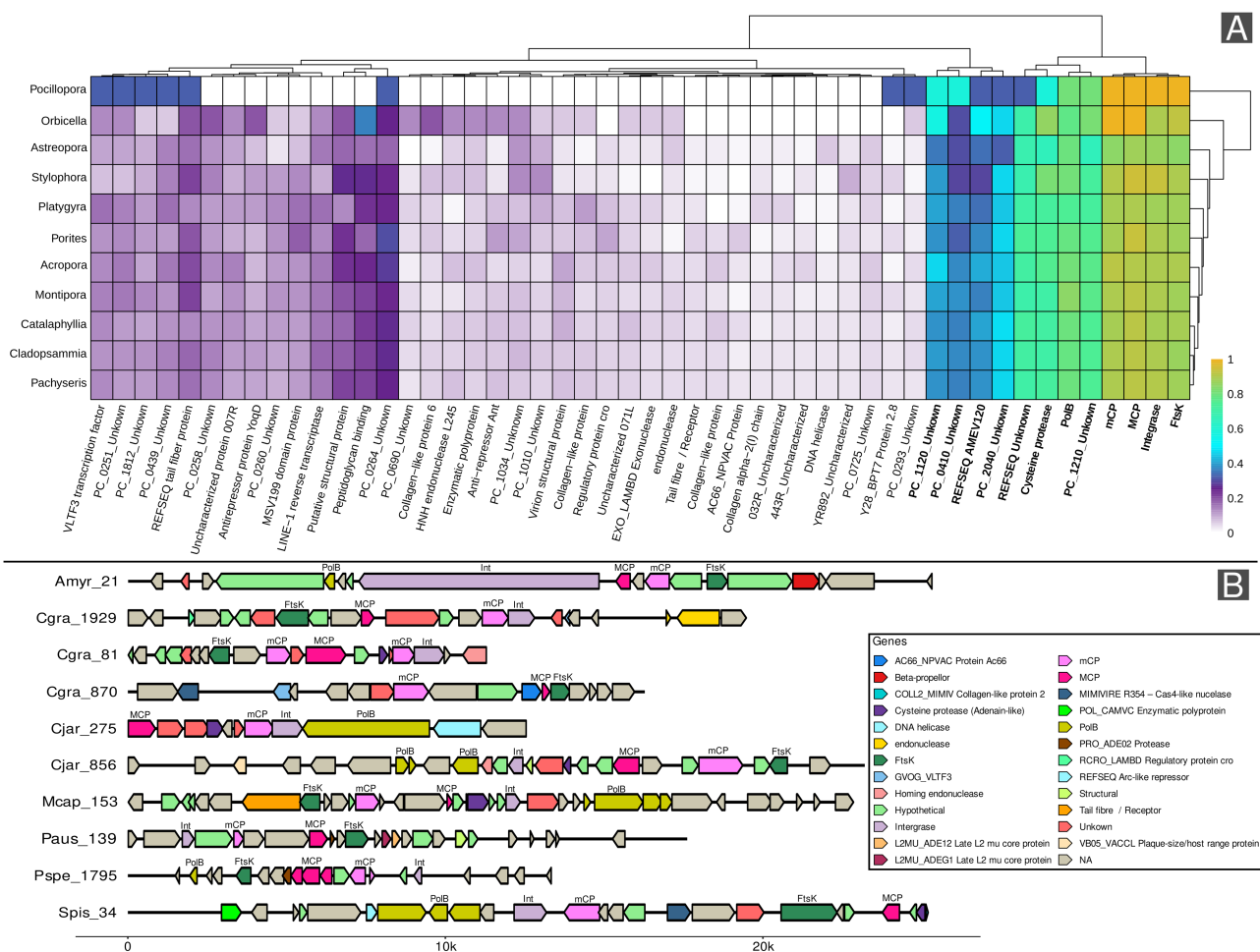


Figure 3. (A) Heatmap of relative proportion of different functions encoded by identified polintovirus EVRs across different genus of stony corals. Each row represents the functional profile of polintovirus in a particular genus, with each column representing annotated proteins occurring in at least 2 per cent of total hits. Hallmark (core) gene labels rendered in bold print. Heatmap was created in the 'pheatmap' R package (Kolde 2019) with clustering performed under euclidean distance similarity measure and 'complete' method for sample clustering in R. (B) Genome structure and gene locations of selected stony coral EVRs. Core genes labeled directly on the map. A selection of EVRs was chosen based on similar length and presence of notable genes (such as VLTf3). The figure was constructed using the gggenomes package in R (Hackl, Ankenbrand, and Adrichem 2023).

presence of many of these genes across these viruses (Fig. 3B) suggests that their genomes (outside of the conserved core genes) were shaped by processes of gene gain and loss. This opens the question on the role of these genes in polintovirus propagation and their impact on host physiology and genome evolution. Recent studies have started to illuminate the accessory gene functions and factors shaping the genomic landscape of polintoviruses. Of note is a recent report on polintoviruses in nematodes, where these elements were found to be a vector of horizontal transfer of key nematode protease and kinase genes (Widen et al. 2023). Specifically, polintoviruses took up these genes as cargoes and transferred them between different nematode species on a global scale. This presents evidence for the potential role of polintoviruses as agents of host genome evolution through the facilitation of horizontal gene transfer and the introduction of novel genetic elements.

Polintovirus gene expression dynamics in *Stylophora pistillata*

Multiple studies have suggested that polintoviruses likely propagate as free virus particles, given the genomic evidence of presence of structural proteins (MCP and mCP) (Krupovic and Koonin 2016;

Koonin and Krupovic 2017; Bellas and Sommaruga 2021). Yet, a clear understanding of the expression dynamics of the genes encoded in these viruses is lacking. In many cases, the presence of a gene is not the direct evidence of its function, as gene expression of mobile genetic elements can be suppressed by host molecular machinery (Almeida et al., 2022; Hollister and Gaut 2009). Although previous studies have managed to identify the expression of PLV MCP from several protist genomes (Bellas et al. 2023; Roitman et al. 2023), a comprehensive assessment of the expression dynamics of the key genes in endogenous polintoviruses remains to be performed.

For an in-depth investigation of the expression patterns of stony coral polintoviruses, we leveraged the data by Rådecker et al. (2021) who sought to identify the transcriptomic response of *S. pistillata* to heat stress utilizing five genetically distinct coral colonies. We identified forty-nine polintovirus EVRs in the *S. pistillata* genome, all of which are distinct at 90 per cent nucleotide identity. By mapping the transcriptomic data from Rådecker et al. (2021) to these elements, we were able to reveal several key aspects of the expression dynamics of polintoviruses. We found that the hallmark genes including MCP, mCP, ATPase, Integrase, and PolB had detectable levels of expression across the samples in both

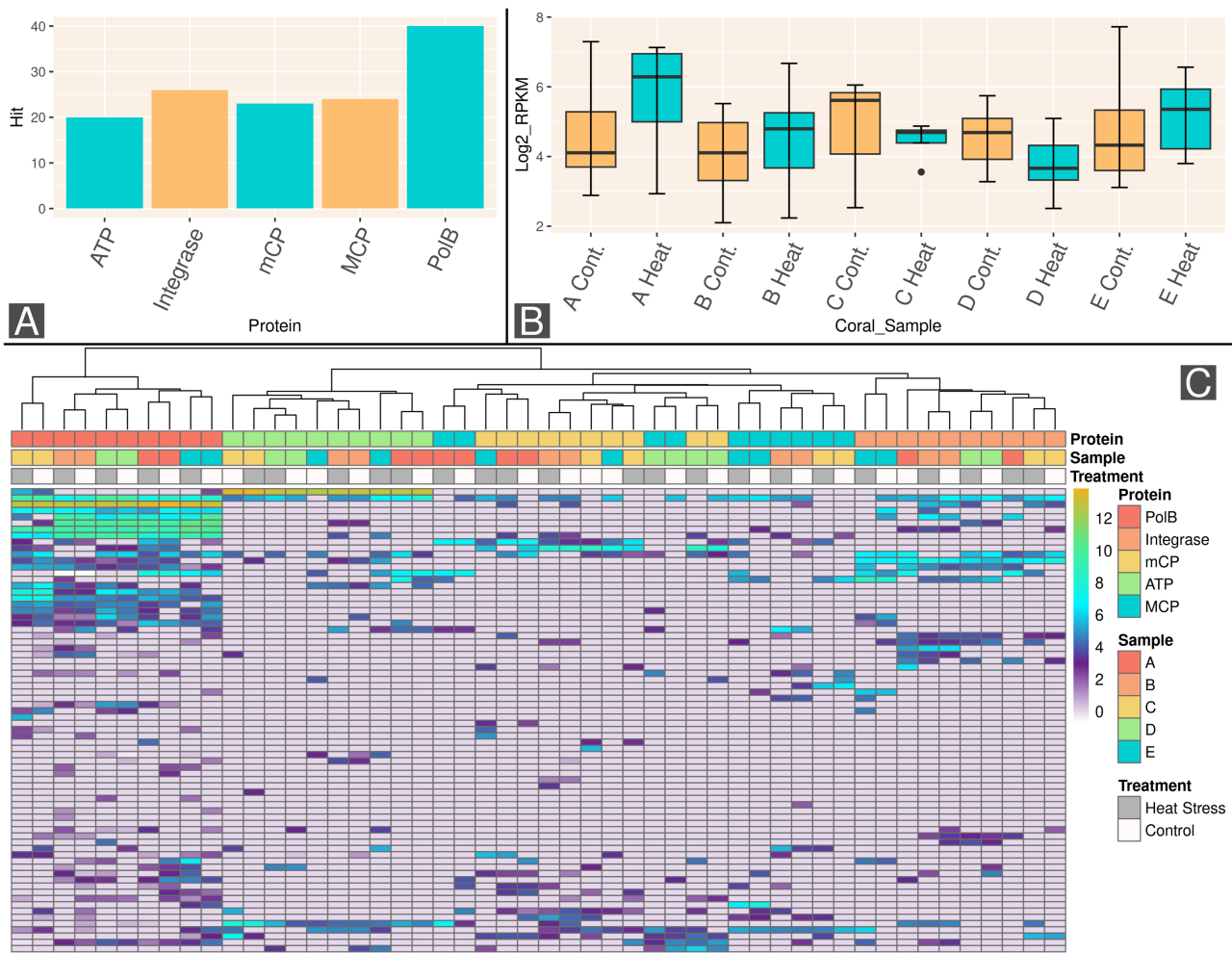


Figure 4. (A) Bar plot of total number of identified *S. pistillata* viral regions (out of forty-nine quality filtered regions) with active polintovirus hallmark gene expression. (B) Boxplot of RPKM for detectable (RPKM >1) MCP expression in *S. pistillata* coral samples. (C) Heatmap of polintovirus core gene expression (PolB, ATPase, Integrase, MCP, and mCP) from identified *S. pistillata* viral regions in genetically distinct coral individuals. Expression levels shown in log2 scale. Annotation layer 1 (top): protein identity. Annotation layer 2: coral sample identity. Annotation layer 3 (bottom): sample treatment type. Each row of the heatmap represents expression data of the core genes from a single EVR. There are forty-nine rows in this heatmap, corresponding to the forty-nine EVRs identified in the *S. pistillata* reference genome. Figure 4C was plotted using the pheatmap (Kolde 2019) package in R. Clustering performed using Euclidean distance measure and ‘complete’ method for sample clustering in R. Columns in heatmap represent identified core proteins from each sample. Rows in heatmap represent individual identified viral regions.

heat stress treated and control samples (Fig. 4A). These genes were expressed in roughly 50 per cent of the forty-nine polintoviruses (MCP = 40 per cent, mCP = 41.67 per cent, ATP = 39.13 per cent, Integrase = 50.75 per cent, PolB = 60.58 per cent) (Supplementary Fig. S11) across different conditions. Because data in this report were generated with genets instead of replicated samples from an individual colony, statistical comparisons on overall levels of polintovirus expression in stressed versus unstressed states cannot be made, but this analysis does provide evidence that all the key genes in these viruses are, in fact, being expressed (Fig. 4B).

Additionally, hierarchical clustering of the hallmark genes expression levels revealed that the expression profile of individual genes across all polintoviruses primarily clustered by gene function (Fig. 4C). Data from PolB, ATPase, and Integrase created gene-specific clusters, whereas the structural proteins (mCP and MCP) formed their own unique cluster; separated from the other genes. This suggests that the transcriptional program of individual polintovirus genes is conserved across coral colonies. If the transcription of the key genes was ‘random’, or result of leaky transcription without an underlying molecular program, they would

be expected to cluster randomly irrespective of gene functions, which is clearly not the case we observed.

Within each of these gene clusters, further subgrouping was driven primarily by samples, not by treatments (Fig. 4C). That is, data from individual colonies clustered together irrespective of stress treatments. Thus, expression of the individual polintovirus genes across different colonies shows colony-specific differences. This phenomenon is similar to well-known biological differences in expression level of the same genes even in different strains of the same microbial species (Pinseel et al. 2022). Therefore, any future studies wishing to examine differences in these expression levels due to various forms of stressors including those relevant to changing ocean conditions (e.g. temperature and pH) should utilize clonal/fragmental replicates to avoid confounding effects introduced by biological or environmental variations among colonies of the same coral species.

The first study reporting polintoviruses, performed by Rinke et al. in 2006 (Rinke and Jurka et al. 2006), proposed a scenario of how polintoviruses propagate within the host genome. In this scenario, the polintovirus element is first excised from the host

genome by the enzyme integrase, leading to an extrachromosomal single-stranded element, which is subsequently replicated by the PolB. After replication, the integrase enzyme catalyzes its integration in the host genome. However, this scenario does not take into account the function of other genes in the polintovirus genome. By utilizing the transcriptomic data, we were able to show that polintoviruses do express the genes involved in capsid formation (MCP and mCP) and genome packaging (ATPase; Fig. 4; Supplementary Fig. S12–16). These genes are only utilized in free living viruses, and thus would likely be lost during genome reduction in case of a strictly endogenous lifestyle. Conservation of these genes and their transcriptomic program supports the possibility that these elements re-enter the environment as free virions as part of their propagation. These scenarios are not mutually exclusive—it is possible that polintoviruses have both life-strategies that include within genome propagation and free particle production. A recent study also supports this possibility where it was shown that polintoviruses have horizontally transferred key genes across nematode lineages through a mechanism that requires free virus production (Widen et al. 2023).

Our analysis of the stony coral viromes (discussed in a previous section) also revealed the presence of Polintovirus MCPs within the virome (Fig. 2). In addition, we have also shown that a group of EVRs in the stony corals cluster within the environmental PLV clade (discussed in the phylogenetics section), members that are known to exist as free virus particles in aquatic ecosystems (Bellas and Sommaruga 2021). Taken together, our findings provide compelling evidence that polintovirus and PLV-like EVRs within stony corals produce free viruses. However, we acknowledge that direct, microscopic observation of free viruses originating from endogenous polintoviruses is yet to be made (Rinke and Jurka et al. 2006; Koonin and Krupovic 2017) and future targeted studies will be required to definitively address this question.

Conclusion

The discovery of a large diversity of polintoviruses and PLV relatives in stony coral genomes and their viromes leads to a number of questions regarding their roles in the ecology, physiology, and genome evolution of this important group of metazoans. The complex evolutionary histories and possible multiple origins of these endogenous elements within the stony coral genomes suggest that they may have originated from previous infections and propagated through vertical transfers and within-genome paralogous expansions. However, future studies focusing on active infections and identification of free polintovirus particles specific to corals will be instrumental in clarifying their origins and integration mechanisms. As these elements seem to have gone through massive lineage-specific expansions in several coral species, our results pose questions regarding their physiological effect and potential for promoting genetic diversity in corals. Particularly relevant is whether the dynamics of these elements could shape the environmental adaptation and response to diverse stressors in stony corals in the global ocean.

Addressing the questions related to the ecological and evolutionary importance of polintoviruses and molecular aspects of their interactions with diverse organisms will require a range of experimental approaches. A key future direction that emerges from our observation is the need to decipher the replication strategy of polintoviruses. Multiple studies have suggested that polintoviruses likely exist as free virus particles (Koonin and Krupovic 2017; Krupovic and Koonin 2017; Bellas and Sommaruga 2021), although direct observation of the particles is yet to be

made (Krupovic, Bamford, and Koonin 2014). A carefully curated clonal replicate experiment at different life-stages of the organism hosting these elements, along with multi-omic analysis at different growth stages in response to diverse stressors, might clarify the replication strategy of these elements. In addition, microscopic observation of free virus particles coupled with targeted gene-knockout approaches would substantiate these observations. Together, the outcome of such experiments could lead to a better understanding of the molecular events underpinning the ecological dynamics of polintoviruses.

Another possible direction is to perform similar genomic analysis of the members of the broader cnidaria phylum using better quality genome assemblies through long read technology. This could help us understand the evolutionary role of these highly prevalent EVEs within this important group of animals crucial for ocean ecosystem health. Resolving the genomic landscape of these elements through long read sequencing would help better quantify the level of expansion of these elements in certain lineages compared to others, along with resolving strain-level heterogeneity in expansion—which might be useful to understand the role of these elements in niche adaptation and species-specific variations in genome evolution.

Polintoviruses and related elements represent an understudied, yet pervasive, driver of genome evolution in diverse eukaryotes ranging from protists to metazoans. Stony corals are a widely studied group of marine animals for which the knowledge base is still being actively developed, given their importance in the stability of ocean ecosystems and response to global change (Madin et al. 2016; Planes and Allemand 2023). Given the widespread prevalence of polintoviruses across the phylogeny of stony corals, they could also serve as important model systems for studying the ecological roles of these viruses and molecular mechanisms that govern their interaction with host organisms.

Materials and methods

Creation of HMM profiles

All proteins identified as belonging to the *Adintoviridae* family of viruses, totaling forty individual species, were downloaded from NCBI (Sayers et al. 2022) (April 2022), and clustered to identify orthologous groups according to a previously reported approach (Aylward et al. 2021). Briefly, we created orthogroups of these *Adintoviridae* proteins using Proteinortho v 6.06 (Lechner et al. 2011) with parameters: `-e 1e-05—identity 25 -p blastp+—selfblast—cov 50 -sim 0.80`. Proteins within each orthogroup were aligned using Clustal Omega (Sievers et al. 2011) and resulting alignments were trimmed using trimal (Capella-Gutierrez, Silla-Martinez, and Gabaldon 2009) (parameter `-gt 0.1`) before conversion to HMM profiles using the `hmmsearch` tool implemented in HMMER (<http://hmmer.org/>) (3.3.2; hmmer.org) with default parameters, resulting in fifty-one distinct orthologous groupings. Similar methods were used to construct HMM profiles from the 2,213 protein clusters (PCs) generated from environmental PLVs reported by Bellas and Sommaruga (2021) which were downloaded around the same time. In addition, we also used several additional HMM profiles specific to the MCPs of virophages and PLVs that were previously constructed and reported by Bellas and Sommaruga (2021) and included with the PCs in their available data.

PLVs are highly diverse and consequently their hallmark genes can show substantial sequence heterogeneity. Due to this, many of the downloaded *Adintoviridae* genes cluster into distinct orthogroups (Bellas and Sommaruga 2021); however, these orthogroups all still correspond to the hallmark genes. For our

analyses related to identification of hallmark genes, we identified all the *Adintoviridae* HMM profiles that correspond to the annotated hallmark genes in the work by [Bellas and Sommaruga \(2021\)](#) and annotated them appropriately for use in our identification process.

Validation of taxonomic assignment of virophages

PCs from the work by [Bellas and Sommaruga \(2021\)](#) were previously annotated by the authors to indicate their identity as PLV or Virophage, and we likewise labeled the *Adintoviridae* as its own distinct group. The taxonomy of the reference genomes was further validated using a recently reported script that can discriminate between virophages and PLV genomes based on gene content similarity ([Roux et al. 2023](#)).

Identification of viral regions

Utilizing the ViralRecall tool ([Aylward and Moniruzzaman 2021](#)) originally developed for the identification of endogenous giant virus signatures in eukaryotic genomes, we identified potential endogenous polintoviruses within thirty coral genomes retrieved from NCBI ([Sayers et al. 2022](#)) in April 2022 (representing the total number of genomes publicly available around that time). To this end, ViralRecall ([Aylward and Moniruzzaman 2021](#)) was tailored so it would identify candidate polintovirus, PLV, and virophage-like regions instead of giant viruses. This was facilitated by replacing the original giant virus orthologous groups HMM profile database that ViralRecall ([Aylward and Moniruzzaman 2021](#)) utilizes for the identification of giant virus regions with the HMM profiles generated using the PCs from the genomes reported by [Bellas and Sommaruga \(2021\)](#) along with the *Adintoviridae* HMM profiles previously created.

For annotation purposes, ViralRecall ([Aylward and Moniruzzaman 2021](#)) outputs the presence of virus-specific hallmark genes as part of the candidate genome identification. In our modified version of ViralRecall ([Aylward and Moniruzzaman 2021](#)), five hallmark genes present in polintoviruses, PLVs, and virophages (MCP, FtsK ATPase, mCP, polB, and integrase) are reported instead of the giant virus-specific hallmark genes.

This modified ViralRecall ([Aylward and Moniruzzaman 2021](#)) program was run on the contigs >10 kb in length from each stony coral genomes with an e-value of <1e-05, a minimum viral region length of 10 kb, and a minimum score of 1. A parameter was used (-fl) to include a 5 kb overhang on each end of the identified viral candidates. All other parameters were set to default.

When modifications were made to the ViralRecall tool, identification thresholds for the new Polintovirus HMMs were relaxed (e-value cutoff 1e-05) to ensure capturing of as many potential viral protein candidates as possible, given the known divergence of many Polintovirus proteins. This served to reduce false-negative rates in our initial screening of these elements.

Resulting viral regions were then manually filtered to exclude regions containing fewer than 1 MCP, 1 mCP, and 1 FtsK ATPase genes (cut-off established based on the genes that are commonly present in all known PLVs, polintoviruses, and virophages ([Koonin and Krupovic 2017](#); [Bellas and Sommaruga 2021](#))). This second step of the semi-automatic pipeline served as a control for false positives within our search efforts. Length cut-off for the regions were then established at a minimum of 10 kb and a maximum size of 50 kb based on the average size of PLV genomes previously reported ([Starrett et al. 2021](#); [Bellas et al. 2023](#)).

Compilation of a MCP database for phylogenetic analysis

The `ncldv_markersearch` script ([Moniruzzaman et al. 2020](#)), originally developed to identify hallmark genes of giant viruses, was modified to incorporate curated hallmark gene HMM profiles specific to PLVs, polintoviruses, and virophages. After modification, `ncldv_markersearch` ([Moniruzzaman et al. 2020](#)) was run on all identified EVR that met established cut-offs. MCP hits obtained by this approach in each species were then clustered at 90 per cent sequence similarity cut-off utilizing CD-Hit ([Fu et al. 2012](#)) (default parameters) to reduce complexity in the dataset while removing redundant sequences for downstream phylogenetic analysis.

Metavirome sequences were downloaded from data provided by [Weynberg et al. \(2017\)](#) and the fastq reads were quality-trimmed using BBduk (sourceforge.net/projects/bbmap/) and assembled using Megahit ([Li et al. 2016](#)) using default parameters. Proteins were predicted in the assembled contigs using prodigal ([Hyatt et al. 2010](#)) with default parameters. The modified `ncldv_markerserach` program ([Moniruzzaman et al. 2020](#)) was then used to identify the MCPs in this dataset as described in the previous paragraph. MCP hits obtained in this manner were also clustered using a 90 per cent similarity cut-off.

[Bellas and Sommaruga \(2021\)](#) compiled a dataset of genomes that contains PLV and virophages identified in their study, as well as PLV and virophage genomes from a number of previous studies and sources ([Bellas, Anesio, and Barker 2015](#); [Yutin et al. 2015](#); [Paez-Espino et al. 2019](#); [Stough et al. 2019](#); [Roux et al. 2023](#)). We used the data from this report as a comprehensive source of references for PLVs and virophages for our phylogenetic analysis. In addition to this, we also included MCPs that were previously reported for *Adintovirus* genomes ([Bellas, Anesio, and Barker 2015](#); [Yutin et al. 2015](#); [Roux et al. 2017](#); [Paez-Espino et al. 2019](#); [Stough et al. 2019](#)) add, which we downloaded from NCBI. We created a final set of MCPs combining the MCP cluster representatives in our study (from stony coral EVRs and metavirome study), along with these aforementioned studies.

Phylogenetic analysis of the MCPs

The compiled dataset of MCPs was aligned using for multiple alignment using fast Fourier transform ([Katoh et al. 2002](#)), and the final alignment was quality trimmed using Trimal ([Capella-Gutierrez, Silla-Martinez, and Gabaldon 2009](#)) (parameter `gt 0.1`). A diagnostic phylogenetic tree using this alignment was then created using Fasttree ([Price, Dehal, and Arkin 2009](#)) which we investigated to manually remove any long branches. The remaining MCP candidates were re-aligned and trimmed utilizing the same process as previously mentioned; and a final maximum likelihood tree was constructed using IQ-Tree v 2 ([Minh et al. 2020](#)) (parameter `-B 1000`; ModelFinder (integrated in the IQ-Tree pipeline) selecting the General Variable Time matrix (VT + R10) model as the best fit), with 1,000 bootstrap iterations ([Hoang et al. 2018](#)) (Fig. 2). The tree was then annotated using iTOL ([Letunic and Bork 2021](#)).

Gene annotation

Functional annotations were performed on the proteins from the identified EVRs in stony coral genomes utilizing the virus orthologous groups (VOG) database (<https://vogdb.org/>), the annotations provided by [Bellas and Sommaruga \(2021\)](#), and our own compiled *Adintoviridae* protein data. To this end, the HMM profiles were searched against proteins from the candidate viral regions in stony coral genomes and significant hits were annotated accordingly.

Identification of TIRs

TIRs flanking the candidate viral regions were located using the IUPACpal program (Alamro et al. 2021). The TIR length of polintoviruses is not defined, and likely varies widely. Thus, for our analysis, minimum and maximum TIR sizes were set at 100 bp and 2,500 bp, respectively, based on the TIR analysis results from Hackl et al. (2021) as a guide. Maximum allowable mismatches between the TIR regions were set to 2 per cent, and the minimum size allowed for the loop was set at 15 kbp (based upon the average size of a standard polintovirus). For IUPACpal (Alamro et al. 2021), the following parameters were used: -M 2500 -m 100 -g 9000.

Statistics

All statistical comparisons were run on R version 4.2.2 (Team, R. C. 2023) using the 'stats' (Team, R. C. 2023) and 'EnvStats' (Millard 2013) packages.

Gene expression profiling of *Stylophora pistillata* polintoviruses

To determine the expression profile of genes encoded by polintoviruses in *S. pistillata*, we downloaded the raw read libraries representing the transcriptome of *S. pistillata* colonies from Rådecker et al. (2021). These data represent five colonies of *S. pistillata* both before and after undergoing heat stress treatment. Raw reads were quality trimmed using trim_galore (Krueger et al. 2023) with default parameters and were mapped to the five core genes (PolB, mCP, MCP, Integrase, FtsK ATPase) of the seventy-four *S. pistillata* polintoviruses using CoverM (<https://github.com/wwood/CoverM>) with a stringent (99 per cent identity) similarity cut-off. RPKM (reads per kilobase pairs per million mapped reads) values were calculated for each of the genes across all ten samples (five control and five heat-treated).

Data availability

The stony coral genomes analyzed in this study are publicly available at NCBI (Sayers et al. 2022) repository (see Supplementary Table S1 for accession numbers). All the identified viral regions, proteins, functional annotations, and raw phylogenetic tree of MCPs are provided in an open access figshare repository (https://figshare.com/articles/dataset/Data_Repository/23620479). The same repository also contains a version of the ViralRecall (Aylward and Moniruzzaman 2021) program that is modified for identification of polintoviruses, PLVs, and Virophages from genome/metagenome sequence data.

Supplementary data

Supplementary data are available at *Virus Evolution* Journal online

Acknowledgements

MM and DS are funded by the Rosenstiel School of Marine, Atmospheric, and Earth Science, University of Miami.

Funding

MM and DS are funded by the Rosenstiel School of Marine, Atmospheric, and Earth Science, University of Miami.

Conflict of interest: The authors declare no conflict of interest.

Author contributions

DS and MM co-conceptualized the study. DS performed the majority of the data analysis, synthesis, and writing. ZF performed part of the graphical data analysis. MM supervised the research and secured funding.

References

- Alamro, H. et al. (2021) 'IUPACpal: Efficient Identification of Inverted Repeats in IUPAC-encoded DNA Sequences', *BMC Bioinformatics*, 22: 51.
- Almeida, M. V. et al. (2022) 'Taming Transposable Elements in Vertebrates: From Epigenetic Silencing to Domestication', *Trends in Genetics*, 38: 529–53.
- Aswad, A., and Katzourakis, A. (2012) 'Paleovirology and Virally Derived Immunity', *Trends in Ecology & Evolution (Amsterdam)*, 27: 627–36.
- Aylward, F. O. et al. (2021) 'A Phylogenomic Framework for Charting the Diversity and Evolution of Giant Viruses', *PLoS Biology*, 19: e3001430.
- Aylward, F. O., and Moniruzzaman, M. (2021) 'ViralRecall—A Flexible Command-Line Tool for the Detection of Giant Virus Signatures in 'Omic Data'', *Viruses*, 13: 150.
- Barreat, J. G. N., and Katzourakis, A. (2023) 'A Billion Years Arms-race between Viruses, Virophages, and Eukaryotes', *eLife*, 12: RP86617.
- Barton, E. S., White, D. W., and Virgin, H. W. (2009) 'Herpesvirus Latency and Symbiotic Protection from Bacterial Infection', *Viral Immunology*, 22: 3–4.
- Bellas, C. et al. (2023) 'Large-scale Invasion of Unicellular Eukaryotic Genomes by Integrating DNA Viruses', *Proceedings of the National Academy of Sciences - PNAS*, 120: e2300465120.
- Bellas, C. M., Anesio, A. M., and Barker, G. (2015) 'Analysis of Virus Genomes from Glacial Environments Reveals Novel Virus Groups with Unusual Host Interactions', *Frontiers in Microbiology*, 6: 656.
- Bellas, C. M., and Sommaruga, R. (2021) 'Polinton-like Viruses are Abundant in Aquatic Ecosystems', *Microbiome*, 9: 13.
- Blanc, G., Gallot-Lavallée, L., and Maumus, F. (2015) 'Proviophages in the *Bigeloviella* Genome Bear Testimony to past Encounters with Giant Viruses', *Proceedings of the National Academy of Sciences - PNAS*, 112: E5318–26.
- Bourque, G. et al. (2018) 'Ten Things You Should Know about Transposable Elements', *Genome Biology*, 19: 199.
- Bushnell B. 6 October, sourceforge.net/projects/bbmap/. 2022. BBMap:BBMap short read aligner, and other bioinformatic tools.
- Calos, M. P., and Miller, J. H. (1980) 'Transposable Elements', *Cell*, 20: 579–595.
- Capella-Gutierrez, S., Silla-Martinez, J. M., and Gabaldon, T. (2009) 'trimAl: A Tool for Automated Alignment Trimming in Large-scale Phylogenetic Analyses', *Bioinformatics*, 25: 1972–3.
- Cróquer, A., and Weil, E. (2009) 'Changes in Caribbean Coral Disease Prevalence after the 2005 Bleaching Event', *Diseases of Aquatic Organisms*, 87: 33–43.
- DAVY, J. E., and Patten, N. L. (2007) 'Morphological Diversity of Virus-like Particles within the Surface Microlayer of Scleractinian Corals', *Aquatic Microbial Ecology: International Journal*, 47: 37–44.
- Dinsdale, E. A. et al. (2008) 'Microbial Ecology of Four Coral Atolls in the Northern Line Islands', *PLoS ONE*, 3: e1584.
- Duerkop, B. A., and Hooper, L. V. (2013) 'Resident Viruses and Their Interactions with the Immune System', *Nature Immunology*, 14: 654–9.

- Feschotte, C., and Pritham, E. J. (2007) 'DNA Transposons and the Evolution of Eukaryotic Genomes', *Annual Review of Genetics*, 41: 331–68.
- Finnegan, D. J. (1992) 'Transposable Elements', *Current Opinion in Genetics & Development*, 2: 861–867.
- Fu, L. et al. (2012) 'CD-HIT: Accelerated for Clustering the Next-generation Sequencing Data', *Bioinformatics*, 28: 3150–2.
- Gardner, T. A. et al. (2003) 'Long-Term Region-Wide Declines in Caribbean Corals', *Science*, 301: 958–60.
- Gilbert, C., Feschotte, C., and Emerman, M. (2010) 'Genomic Fossils Calibrate the Long-Term Evolution of Hepadnaviruses', *PLoS Biology*, 8: e1000495.
- Grasis, J. A. et al. (2017) 'The Intra-Dependence of Viruses and the Holobiont', *Frontiers in Immunology*, 8: 1501.
- Green, E. P., and Bruckner, A. W. (2000) 'The Significance of Coral Disease Epizootiology for Coral Reef Conservation', *Biological Conservation*, 96: 347–61.
- Haapa-Paananen, S., Wahlberg, N., and Savilahti, H. (2014) 'Phylogenetic Analysis of Maverick/Polinton Giant Transposons across Organisms', *Molecular Phylogenetics & Evolution*, 78: 271–4.
- Hackl, T. et al. (2021) 'Virophages and Retrotransposons Colonize the Genomes of a Heterotrophic Flagellate', *bioRxiv*, 10: e27674.
- Hackl, T., Ankenbrand, M. J., and Adrichem, B. V. 'Gggenomes: A Grammar of Graphics for Comparative Genomics', 2023.
- Hoang, D. T. et al. (2018) 'UFBoot2: Improving the Ultrafast Bootstrap Approximation', *Molecular Biology and Evolution*, 35: 518–22.
- Hoegh-Guldberg, O. et al. (2007) 'Coral Reefs under Rapid Climate Change and Ocean Acidification', *refScience*, 318: 1737–42.
- Hollister, J. D., and Gaut, B. S. (2009) 'Epigenetic Silencing of Transposable Elements: A Trade-off between Reduced Transposition and Deleterious Effects on Neighboring Gene Expression', *Genome Research*, 19: 1419–28.
- Holmes, E. (2011) 'The Evolution of Endogenous Viral Elements', *Cell Host & Microbe*, 10: 368–77.
- Howard Hughes Medical Institute (2015), HMMER 3.1b2 <<http://hmmer.org/>> accessed February 2015.
- Howe-Kerr, L. et al. (2022) 'Viruses of a Key Coral Symbiont Exhibit Temperature-driven Productivity across a Reefscape', *ISME COMMUN*, 3: 27.
- Hughes, T. P. et al. (2003) 'Climate Change, Human Impacts, and the Resilience of Coral Reefs', *Science*, 301: 929–33.
- Hyatt, D. et al. (2010) 'Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification', *BMC Bioinformatics*, 11: 119.
- Jeong, D. et al. (2023) 'DNA Polymerase Diversity Reveals Multiple Incursions of Polintons During Nematode Evolution', *bioRxiv*, 40: msad274.
- Katoh, K. et al. (2002) 'MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform', *Nucleic Acids Research*, 30: 3059–66.
- Katzourakis, A., Gifford, R. J., and Malik, H. S. (2010) 'Endogenous Viral Elements in Animal Genomes', *PLoS Genetics*, 6: e1001191.
- Knowlton, N. (2001) 'The Future of Coral Reefs', *Proceedings of the National Academy of Sciences - PNAS*, 98: 5419–25.
- Knowlton, N., and Rohwer, F. (2003) 'Multispecies Microbial Mutualisms on Coral Reefs: The Host as a Habitat', *The American Naturalist*, 162: S51–S62.
- Kolde, R. *Pheatmap: Pretty Heatmaps*, 2019.
- Koonin, E. V., and Krupovic, M. (2017) 'Polintons, Virophages and Transpovirons: A Tangled Web Linking Viruses, Transposons and Immunity', *Current Opinion in Virology*, 25: 7–15.
- Felix, K. et al. (2023). FelixKrueger/TrimGalore: v0.6.10 - add default decompression path (0.6.10). Zenodo.
- Krupovic, M., Bamford, D. H., and Koonin, E. V. (2014) 'Conservation of Major and Minor Jelly-roll Capsid Proteins in Polinton (Maverick) Transposons Suggests that They are Bona Fide Viruses', *Biology Direct*, 9: 6.
- Krupovic, M., and Koonin, E. V. (2016) 'Self-synthesizing Transposons: Unexpected Key Players in the Evolution of Viruses and Defense Systems', *Current Opinion in Microbiology*, 31: 25–33.
- Lajeunesse, T. C. et al. (2018) 'Systematic Revision of Symbiodiniaceae Highlights the Antiquity and Diversity of Coral Endosymbionts', *Current Biology*, 28: 2570–2580.e6.
- Lechner, M. et al. (2011) 'Proteinortho: Detection of (Co-)orthologs in Large-scale Analysis', *BMC Bioinformatics*, 12: 124.
- Letunic, I., and Bork, P. (2021) 'Interactive Tree of Life (ItoL) V5: An Online Tool for Phylogenetic Tree Display and Annotation', *Nucleic Acids Research*, 49: W293–6.
- Li, D. et al. (2016) 'MEGAHIT V1.0: A Fast and Scalable Metagenome Assembler Driven by Advanced Methodologies and Community Practices', *Methods San Diego Calif*, 102: 3–11.
- Liu, H. et al. (2011) 'Widespread Horizontal Gene Transfer from Circular Single-stranded DNA Viruses to Eukaryotic Genomes', *BMC Evolutionary Biology*, 11: 276.
- Loch, J. A. et al. (2008) 'Accurate Whole Human Genome Sequencing Using Reversible Terminator Chemistry', *Nature*, 456: 53–9.
- Lombard, F. et al. (2023) 'Open Science Resources from the Tara Pacific Expedition across Coral Reef and Surface Ocean Ecosystems', *Scientific Data*, 10: 324.
- Madin, J. S. et al. (2016) 'The Coral Trait Database, a Curated Database of Trait Information for Coral Species from the Global Oceans', *Scientific Data*, 3: 160017.
- Matthews, J. L. et al. (2020) 'Symbiodiniaceae-bacteria Interactions: Rethinking Metabolite Exchange in Reef-building Corals as Multi-partner Metabolic Networks', *Environmental Microbiology*, 22: 1675–87.
- Mera, H., and Bourne, D. G. (2017) 'Disentangling Causation: Complex Roles of Coral-associated Microorganisms in Disease', *Environ Microbiol*, 20: 431–449.
- Millard, S. P. (2013) *EnvStats: An R Package for Environmental Statistics*. New York, NY, USA: Springer, 2nd edn.
- Minh, B. Q. et al. (2020) 'IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era', *Molecular Biology and Evolution*, 37: 1530–4.
- Moniruzzaman, M. et al. (2020) 'Dynamic Genome Evolution and Complex Virocell Metabolism of Globally-distributed Giant Viruses', *Nature Communications*, 11: 1710.
- Mougari, S. et al. (2020) 'A Virophage Cross-species Infection through Mutant Selection Represses Giant Virus Propagation, Promoting Host Cell Survival', *Communications Biology*, 3: 248.
- Muller, E. M., and van Woesik, R. (2012) 'Caribbean Coral Diseases: Primary Transmission or Secondary Infection? *Global Change Biology*', 18: 3529–35.
- (2017) 'Multiple Origins of Viral Capsid Proteins from Cellular Ancestors', *Proceedings of the National Academy of Sciences - PNAS*, 114: E2401–10.
- Munoz-Lopez, M., and Garcia-Perez, J. (2010) 'DNA Transposons: Nature and Applications in Genomics', *Current Genomics*, 11: 115–28.
- Noel, B. et al. (2023) 'Pervasive Tandem Duplications and Convergent Evolution Shape Coral Genomes', *Genome Biology*, 24: 123.
- Paez-Espino, D. et al. (2019) 'Diversity, Evolution, and Classification of Virophages Uncovered through Global Metagenomics', *Microbiome*, 7: 157.
- Pandolfi, J. M. et al. (2003) 'Global Trajectories of the Long-Term Decline of Coral Reef Ecosystems', *Science*, 301: 955–8.

- Patel, M. R., Emerman, M., and Malik, H. S. (2011) 'Paleovirology — Ghosts and Gifts of Viruses Past', *Current Opinion in Virology*, 1: 304–9.
- Pinseel, E. et al. (2022) 'Strain-specific Transcriptional Responses Overshadow Salinity Effects in a Marine Diatom Sampled along the Baltic Sea Salinity Cline', *The ISME Journal*, 16: 1776–87.
- Planes, S., and Allemand, D. (2023) 'Insights and Achievements from the Tara Pacific Expedition', *Nature Communications*, 14: 3131.
- Pollock, F. J. et al. (2011) 'The Urgent Need for Robust Coral Disease Diagnostics', *PLoS Pathogens*, 7: e1002183.
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2009) 'FastTree: Computing Large Minimum Evolution Trees with Profiles Instead of a Distance Matrix', *Molecular Biology and Evolution*, 26: 1641–50.
- Rädecker, N. et al. (2021) 'Heat Stress Destabilizes Symbiotic Nutrient Cycling in Corals', *Proceedings of the National Academy of Sciences of the United States of America*, 118: e2022653118.
- (2004) 'Black Band Disease', *Coral Health and Disease*, 325–36.
- Richardson, L. L. (1998) 'Coral Diseases: What Is Really Known', *Trends in Ecology and Evolution*, 13: 438–43.
- Rinke, E. M. et al. (2006) 'Self-synthesizing DNA Transposons in Eukaryotes', *Proceedings of the National Academy of Sciences of the United States of America*, 103: 4540–5.
- Roitman, S. et al. (2023) 'Isolation and Infection Cycle of a Polinton-like Virus Virophage in an Abundant Marine Alga', *Nature Microbiology*, 8: 332–46.
- Roux, S. et al. (2017) 'Ecogenomics of Virophages and Their Giant Virus Hosts Assessed through Time Series Metagenomics', *Nature Communications*, 8: 858–12.
- Roux, S. et al. (2023) 'Updated Virophage Taxonomy and Distinction from Polinton-like Viruses', *Biomolecules (Basel, Switzerland)*, 13: 204.
- Sayers, E. W. et al. (2022) 'Database Resources of the National Center for Biotechnology Information', *Nucleic Acids Research*, 50: D20–6.
- Sievers, F. et al. (2011) 'Fast, Scalable Generation of High-quality Protein Multiple Sequence Alignments Using Clustal Omega', *Molecular Systems Biology*, 7: 539–n/a.
- Souter, P. et al. 'Status of Coral Reefs of the World: 2020', *Global Coral Reef Monitoring Network (GCRMN), International Coral Reef Initiative (ICRI)*, 2021.
- Starrett, G. J. et al. (2021) 'Adintoviruses: A Proposed Animal-tropic Family of Midsize Eukaryotic Linear dsDNA (MELD) Viruses', *Virus Evolution*, 7: veaa055.
- Stough, J. M. A. et al. (2019) 'Genome and Environmental Activity of a Chrysochromulina Parva Virus and Its Virophages', *Frontiers in Microbiology*, 10: 703.
- Team, R. C.. R: A Language and Environment for Statistical Computing R Foundation for Statistical Computing, Vienna, Austria, 2023.
- Thurber, R. V. et al. (2009) 'Metagenomic Analysis of Stressed Coral Holobionts', *Environmental Microbiology*, 11: 2148–63.
- Thurber, R. V. et al. (2017) 'Virus–host Interactions and Their Roles in Coral Reef Health and Disease', *Nature Reviews, Microbiology*, 15: 205–16.
- University of Wien. (2023), VOGDB Virus Orthologous Groups <<https://vogdb.org/>> accessed 22 June.
- Volkman, H. E., and Stetson, D. B. (2014) 'The Enemy Within: Endogenous Retroelements and Autoimmune Disease', *NATURE IMMUNOLOGY*, 15: 415–22.
- Walker, P. J. et al. (2021) 'Changes to Virus Taxonomy and to the International Code of Virus Classification and Nomenclature Ratified by the International Committee on Taxonomy of Viruses (2021)', *Archives of Virology*, 166: 2633–48.
- Webster, N. S. et al. (2016) 'Host-associated Coral Reef Microbes Respond to the Cumulative Pressures of Ocean Warming and Ocean Acidification', *Scientific Reports*, 6: 19324.
- Weynberg, K. D. et al. (2017) 'Coral-associated Viral Communities Show High Levels of Diversity and Host Auxiliary Functions', *PeerJ*, 5: e4054v.
- Widen, S. A. et al. (2023) 'Virus-like Transposons Cross the Species Barrier and Drive the Evolution of Genetic Incompatibilities', *Science (American Association for the Advancement of Science)*, 380: eade0705.
- Wiersinga, W. J. et al. (2014) 'Host Innate Immune Responses to Sepsis', *Virulence*, 5: 36–44.
- Wilson, W. H. et al. (2005) 'An Enemy Within? Observations of Virus-like Particles in Reef Corals', *Coral Reefs*, 24: 145–8.
- Woesik, R., and Randall, C. J. (2017) 'Coral Disease Hotspots in the Caribbean', *Ecosphere (Washington, D.C.)*, 8: e01814–n/a.
- Woodcroft, B. Queensland University of Technology (2021), *CoverM* <<https://github.com/wwood/CoverM>> accessed 24 February.
- Worobey, M. et al. (2010) 'Island Biogeography Reveals the Deep History of SIV', *Science (American Association for the Advancement of Science)*, 329: 1487.
- Yutin, N. et al. (2015) 'A Novel Group of Diverse Polinton-like Viruses Discovered by Metagenome Analysis', *BMC Biology*, 13.
- Yutin, N., Raoult, D., and Koonin, E. V. (2013) 'Virophages, Polintons, and Transpovirons: A Complex Evolutionary Network of Diverse Selfish Genetic Elements with Different Reproduction Strategies', *Virology Journal*, 10: 158.
- Zhuo, X., Feschotte, C., and Belshaw, R. (2015) 'Cross-Species Transmission and Differential Fate of an Endogenous Retrovirus in Three Mammal Lineages', *PLoS Pathogens*, 11: e1005279.