bioRxiv preprint doi: https://doi.org/10.1101/2022.04.11.487905; this version posted April 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Endogenous viral elements reveal associations between a non-retroviral RNA virus and symbiotic dinoflagellate genomes

Alex J. Veglia[&1], Kalia S.I. Bistolas[&2]*, Christian R, Voolstra[3], Benjamin C. C. Hume[3], Serge Planes[4], Denis Allemand[5], Emilie Boissin[4], Patrick Wincker[6], Julie Poulain[6], Clémentine Moulin[7], Guillaume Bourdin[8], Guillaume Iwankow[4], Sarah Romac[9], Sylvain Agostini10, Bernard Banaigs[4], Emmanuel Boss[8], Chris Bowler[11], Colomban de Vargas[9], Eric Douville[12], Michel Flores[13], Didier Forcioli[14], Paola Furla[14], Pierre Galand[15], Eric Gilson[14], Fabien Lombard[16], Stéphane Pesant[17], Stéphanie Reynaud[18], Shinichi Sunagawa[19], Olivier Thomas[20], Romain Troublé[7], Didier Zoccola[18], Adrienne M.S. Correa[1], and Rebecca L. Vega Thurber[2]

[1] BioSciences Department, Rice University, Houston, TX, USA
[2] Microbiology Department, Oregon State University, Corvallis, OR, USA
[3] Department of Biology, University of Konstanz, Konstanz, Germany
[4] PSL Research University: EPHE-UPVD-CNRS, USR 3278 CRIOBE, Laboratoire d'Excellence CORAIL, Université de Perpignan, 52 Avenue Paul Alduy, 66860 Perpignan Cedex, France
[5] Centre Scientifique de Monaco, 8 Quai Antoine Ier, MC-98000, Principality of Monaco
[6] Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, Evry, France
[7] Fondation Tara Océan, Base Tara, 8 rue de Prague, 75 012 Paris, France
[8] School of Marine Sciences, University of Maine, USA
[9] Sorbonne Université, CNRS, Station Biologique de Roscoff, AD2M, UMR 7144, ECOMAP, Roscoff, France
[10] Shimoda Marine Research Center, University of Tsukuba, 5-10-1, Shimoda, Shizuoka, Japan
[11] Institut de Biologie de l'Ecole Normale Supérieure (IBENS), Ecole normale supérieure, CNRS, INSERM, Université PSL, 75005 Paris, France
[12] Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France
[13] Weizmann Institute of Science, Department of Earth and Planetary Sciences, 76100 Rehovot, Israel
[14] Université Côte d'Azur, CNRS, INSERM, IRCAN, Medical School, Nice, France and Department of Medical Genetics, CHU of Nice, France
[15] Sorbonne Université, CNRS, Laboratoire d'Ecogéochimie des Environnements Benthiques (LECOB), Observatoire Océanologique de Banyuls, 66650 Banyuls sur mer, France
[16] Sorbonne Université, Institut de la Mer de Villefranche sur mer, Laboratoire d'Océanographie de Villefranche, F-06230 Villefranche-sur-Mer, France
[17] European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK
[18] Centre Scientifique de Monaco, 8 Quai Antoine Ier, MC-98000, Principality of Monaco
[19] Department of Biology, Institute of Microbiology and Swiss Institute of Bioinformatics, Vladimir-Prelog-Weg 4, ETH Zürich, CH-8093 Zürich, Switzerland
[20] Marine Biodiscovery Laboratory, School of Chemistry and Ryan Institute, National University of Ireland, Galway, Ireland

***Corresponding Author: Dr. Kalia Bistolas** ksb97@cornell.edu, 415 Nash Hall, Oregon State University, Corvallis, OR 97331; orcid: 0000-0002-1664-5028
Contributions [&]Authors contributed equally

Running head: Endogenization of a non-retroviral RNA virus in Symbiodiniaceae genomes

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.11.487905; this version posted April 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Keywords: coral reef, endogenous viral element (EVE), genome, RNA virus, Symbiodiniaceae, symbiosis

Abstract

Endogenous viral elements (EVEs) offer insight into the evolutionary histories and hosts of contemporary viruses. This study leveraged DNA metagenomics and genomics to detect and infer the host of a non-retroviral dinoflagellate-infecting +ssRNA virus (dinoRNAV) common in coral reefs. As part of the Tara Pacific Expedition, this study surveyed 269 newly sequenced cnidarians and their resident symbiotic dinoflagellates (Symbiodiniaceae), associated metabarcodes, and publicly available metagenomes, revealing 178 dinoRNAV EVEs, predominantly among hydrocoral-dinoflagellate metagenomes. Putative associations between Symbiodiniaceae and dinoRNAV EVEs were corroborated by the characterization of dinoRNAV-like sequences in 17 of 18 scaffold-scale and one chromosome-scale dinoflagellate genome assembly, flanked by characteristically cellular sequences and in proximity to retroelements, suggesting potential mechanisms of integration. EVEs were not detected in dinoflagellate-free (aposymbiotic) cnidarian genome assemblies, including stony corals, hydrocorals, jellyfish, or seawater. The pervasive nature of dinoRNAV EVEs within dinoflagellate genomes (especially *Symbiodinium*), as well as their inconsistent within-genome distribution and fragmented nature, suggest ancestral or recurrent integration of this virus with variable conservation. Broadly, these findings illustrate how +ssRNA viruses may obscure their genomes as members of nested symbioses, with implications for host evolution, exaptation, and immunity in the context of reef health and disease.

Introduction

Endogenous viral elements, or "EVEs," arise when whole or fragmented viral genomes are incorporated into host cell germlines. Once integrated, EVEs may propagate across successive host generations, potentially becoming fixed in a population through natural selection or drift (Johnson 2015, 2019). Therefore, the presence and content of EVEs can provide clues into the evolutionary relationships among host species and shed light on ancient and modern virus-host interactions (Johnson 2010). To date, most EVEs described in metazoan and plant genomes are retroviral, as this viral group must integrate their genome (as a provirus) into the genome of the host to replicate. Retroviruses thus possess and encode all of the molecular machinery (e.g. reverse transcriptases, integrases) required to integrate autonomously (Stoye 2012). Remarkably, however, sequences from viruses that do not encode reverse transcriptases or exploit integration as a component of an obligate replication strategy – even viruses with no DNA stage – have also recently been detected as EVEs in diverse eukaryotic genomes (Gallot-Lavallée & Blanc 2017, Flynn and Moreau 2019, Horie et al. 2010, Katzourakis & Gifford 2010, Chiba et al. 2011, Chu et al. 2014, Kojima et al. 2021). These non-retroviral RNA EVEs have been reported in hosts ranging from unicellular algae to chiropteran (bat) genomes (Ballinger et al. 2012, Tromas et al. 2014, Palantini et al. 2017, Wang et al. 2014, Jebb et al. 2020, Moniruzzaman et al. 2020, Skirmuntt et al. 2020). Though the mechanisms behind non-retroviral integration continue to be explored, viral sequences may be introduced via nonhomologous recombination and repair, through interactions with host-provisioned integrases and reverse transcriptases supplied on mobile elements (e.g. retrotransposons), or by utilizing co-infecting viruses (Horie et al. 2010, Flynn & Moreau 2019).

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.11.487905; this version posted April 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Endogenization of any viral sequence (including non-retroviral EVEs) may have positive, neutral or negative effects on a host (Roossinck 2011, Harrison & Brockhurst 2017, Correa et al. 2021). While many EVEs are functionally defective or deleterious and ultimately removed from a population via purifying selection, retained EVEs may remodel the genomic architecture of their hosts or introduce sources of genetic innovation later co-opted for host function (i.e. exaptation; Jern & Coffin 2008, Oliveira et al. 2008). Such 'domesticated' EVEs are utilized by hosts as regulatory elements, transcription factors, transposons, and templates for protein coding functions for purposes ranging from developmental pathways to brain function (Feschotte & Gilbert 2012, Frank & Feschotte 2017, Sofuku & Honda 2017, Takahashi et al. 2019). In particular, non-retroviral EVEs potentially serve as antiviral prototypes that help hosts combat infection by exogenous viruses currently circulating in the population (e.g. RNAi; Witfield et al. 2017, Ter Horst 2019, Palantini et al. 2017, Suzuki et al. 2020). If expressed, EVEs may have a significant influence on the health, physiology and/or behavior of their hosts in natural and experimental systems (Parker & Brisson 2019, Suzuki et al. 2020, Wilson et al. 2001).

Investigating the distribution, sequence identity, and function of EVEs can yield insight into virus-host interactions across generations. EVEs catalogue a subset of the viruses that a host lineage has encountered and can link homologous extant viruses to contemporary hosts or known disease states (Holmes 2011, Suzuki et al. 2020). Because integrated elements may accrue mutations at a slower rate than exogenous viral genomes (Aiewsakun and Katzourakis, 2015, Flynn & Moreau 2019), EVEs can fill gaps in virus-host networks and act as synapomorphies, indicating the minimum time that a virus may have interacted with a host. As 'genomic fossils', EVEs have helped paleovirologists date the minimum origin of *Circoviridae, Hepadnaviridae, Bornaviridae, Orthornavirae, Lentiviridae*, and *Spumaviridae* infections within metazoans (Feschotte & Gilbert 2012, Johnson 2019). The evolutionary and ecological contexts that EVEs provide for exogenous viruses are particularly informative in understanding the interactions of viruses with multipartite or nested symbiotic systems (Patel et al. 2011, Katzourakis 2013, Aiewsakun & Katzourakis, 2015).

Coral holobionts – the cnidarian animal and its resident microbial assemblage, including dinoflagellates in the family Symbiodiniaceae, bacteria, archaea, fungi, and viruses – are an ecologically and economically valuable, multipartite non-model system (Knowlton & Rohwer, 2003, Matthews et al. 2020). Symbiodiniaceae are key obligate nutritional symbionts of corals and support their hosts in the construction of reef frameworks (LaJeunesse et al. 2018). However, environmental stress can break down coral-Symbiodiniaceae partnerships, resulting in bleaching – the mass loss of Symbiodiniaceae cells (Glynn 1996). Some bleaching signs (paling of a coral colony) are hypothesized to also result from viral lysis of Symbiodiniaceae (van Oppen et al. 2009, Correa et al. 2016, Vega Thurber et al. 2017, Messyasz et al. 2020, Correa et al. 2021, Grupstra et al. 2022), but direct evidence supporting this hypothesis remains limited. Overall, the role of viruses in coral colony health and disease requires further examination.

Non-retroviral +ssRNA dinoRNAV sequences were first reported in stony corals based on five metatranscriptomic sequences and corroborated by Symbiodiniaceae EST libraries (Correa et al. 2013). Subsequent studies indicated that similar +ssRNA viruses are commonly detected in coral RNA viromes and metatranscriptomes, as well as via targeted amplicon assays (Weynberg et al. 2014, Levin et al. 2017, Montalvo-Proano et al. 2017, Grupstra et al. 2022). These viruses exhibit synteny and significant homology to Heterocapsa circularisquama RNA virus (HcRNAV; Levin et al. 2017), the sole recognized representative of the genus *Dinornavirus* and a known pathogen of free-living dinoflagellates (Nagasaki et al. 2005). Both

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.11.487905; this version posted April 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

HcRNAV and dinoRNAV sequences detected in coral holobiont tissues contain two ORFs – a Major Capsid Protein (*MCP)* and RNA dependent RNA polymerase (*RdRp*). Furthermore, icosahedral virus-like particle (VLP) arrays resembling HcRNAV (but with 40% smaller individual particle diameters) have been imaged in the Symbiodiniaceae-dense coral gastrodermis tissue and in Symbiodiniaceae themselves (Lawrence et al. 2014). Levin et al. (2017) assembled the 5.2kb genome of a putative dinoRNAV from a poly(A)-selected metatranscriptome generated from cultured *Symbiodinium*. The assembly contained a 5' dinoflagellate spliced leader ("dinoSL"; Zhang et al. 2013) — a component of >95% of Symbiodiniaceae mRNAs, speculated to illustrate molecular mimicry — and exhibited >1000-fold higher expression in a thermosensitive *Cladocopium C1* population relative to a thermotolerant population of this Symbiodiniaceae strain at ambient temperatures (27°C, Levin et al. 2017, LaJeunesse et al. 2018). Together, the findings from these studies suggest that Symbiodiniaceae are target hosts of reef-associated dinoRNAVs.

This study (1) systematically searched for putative endogenized dinoRNAVs in metagenomes from *in situ* (symbiotic) coral colonies and seawater, as well as in available genomes of Symbiodiniaceae and aposymbiotic (symbiont-free) cnidarians, (2) investigated the evolutionary relationship of putative dinoRNAV EVEs to exogenous reef-associated dinoRNAV sequences, and (3) made preliminary inferences regarding the distribution and possible function of these dinoRNAV EVEs based on their detection, prevalence, and genomic context.

Methods

Identification and computational validation of dinoRNAV EVEs leveraging meta'omics

The Tara Pacific Expedition (2016-2018) sampled coral reefs to investigate reef health and ecology using multiple methods, including amplicon sequencing and metagenomics (Planes et al. 2019). In this study, we explored metagenomes generated from hydrocorals (n=60 *Millepora*), stony corals (n=108 *Porites*, n=101 *Pocillopora*) sampled from 11 islands across the South Pacific Ocean during the Tara Pacific Expedition for dinoRNAV EVEs (Figure 1, Supplementary Table 1A, 1B; Pesant et al. 2020). Amplicon libraries of the dinoflagellate Internal Transcribed Spacer 2 (ITS2) gene fragment were sequenced in tandem with the metagenomes, to characterize the dominant Symbiodiniaceae harbored by hydrozoan and stony coral colonies (Hume et al. 2020).

To confirm that these dinoRNAV EVE sequences were affiliated with coral holobionts and reduce the possibility that they are technical artifacts, publicly available metagenome libraries were analyzed. These additional libraries included 120 assembled pelagic water samples presumed to include pelagic dinoflagellate sequences from the Tara Oceans dataset (2009-2013; Pesant et al. 2015) and 30 MiSeq metagenomes from unfractionated samples of the stony coral genus *Acropora*, which were processed and sequenced via a different pipeline (Supplemental Table 1B, Supplemental Figure 1). Publicly accessible transcriptomes from *Symbiodinium microadriaticum* (Supplemental Table 1B) were also queried to determine if dinoRNAV-like sequences were present in poly(A)-selected dinoflagellate transcriptomes and resembled EVEs in terms of proximal gene composition and presence of a characteristic pre-mRNA spliced leader (SL) sequence (as in Levin et al, 2017). Details regarding the collection of samples, generation of metagenomes and associated Symbiodiniaceae amplicon libraries, and associated bioinformatic analyses are provided in Supplementary Figure 1).

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.11.487905; this version posted April 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Metagenomic and transcriptomic scaffolds were annotated against a curated database of dinoRNAV-like sequences (Supplemental Table 2) via BLASTx (Altschul et al. 1990, Supplementary Figure 1). Alignments to the custom database with a bit score <50 and percent shared amino acid identity <30% were excluded from further analysis. A length penalty was not imposed during this step due to the limited length of assembled scaffolds (average N50=3341±127 nt across all queried libraries). Open reading frames (ORFs) from selected scaffolds were called via Prodigal (v.2.6.3; Hyatt et al. 2010) and annotated against the NCBI-nr database (DIAMOND v.2.0.6; Buchfink et al. 2015) to confirm homology to dinoRNAVs and identify adjacent dinoflagellate sequences. In the absence of complete ORFs (potentially due to the limited size of scaffolds, partial integrations, etc.), homology was confirmed through comparison of the initial alignments to the curated database and 300nt of upstream/downstream flanking sequences (bedtools v.2.30.0; Quinlan et al, 2010) against the NCBI-nr database. Non-normalized quality-controlled reads were mapped via bbmap (v.38.84; Bushnell et al. 2017), and putative EVEs were assessed for uniform read coverage across scaffolds, reducing the likelihood of chimeric assembly. RNA secondary structure was predicted via mfold (v.3.5; Zuker et al 2003).

dinoRNAV EVEs in dinoflagellate and aposymbiotic cnidarian genomes

Publicly available dinoflagellate and aposymbiotic (dinoflagellate-free) cnidarian genome assemblies were queried to resolve the putative host(s) of dinoRNAVs, to assess homology among detected dinoRNAVs within coral holobionts, and to compare genes proximal to dinoRNAV EVEs in different host species/strains. A chromosome-scale dinoflagellate genome assembly generated from a *Symbiodinium microadriaticum* culture (Accession: GSE152150, Nand et al. 2021), and scaffold-scale genome assemblies were examined for dinoRNAV EVEs (Supplemental Table 1B). Scaffold-scale genome assemblies were from the closely related families Symbiodiniaceae and Suessiaceae, and included representatives from the genera *Symbiodinium* (n=9), *Breviolum* (n=1), *Cladocopium* (n=3), *Durusdinium* (n=1), *Fugacium* (n=2), and *Polarella* (n=2), as well as 25 aposymbiotic cnidarian genome assemblies, including the stony coral genera *Acropora* (n=13), *Astreopora* (n=1), *Galaxea* (n=1), *Montastraea* (n=1), *Montipora* (n=3), *Orbicella* (n=1), *Pocillopora* (n=2), *Porites* (n=1), and *Stylophora* (n=1), and the jellyfish *Clytia* (n=1; Figure 2, Supplemental Table 1B). Genome completeness and quality were assessed via BUSCO (v3; Simão et al. 2015) with the Eukaryota dataset and QUAST (v5.0.2; Gurevich et al. 2013), respectively. Scaffolds/chromosomes containing putative dinoRNAV EVEs were identified by aligning sequences to the protein version of the Reference Viral DataBase (RVDB v.19; Bigot et al. 2019) using DIAMOND BLASTx (v0.9.30; Buchfink et al. 2015). The same exclusion criteria were maintained for alignments of metagenomic scaffolds, also omitting alignments <100 amino acids. Regions of dinoflagellate genomes exhibiting similarity to the *MCP* or *RdRp* of reef-associated dinoRNAV reference genomes (Levin et al. 2017) or other closely related +ssRNA viruses (Supplemental Table 2) were extracted and re-aligned to the NCBI-nr database to further confirm viral homology. Putative whole dinoRNAV-like genomes within scaffolds were identified based on the presence of *MCP* and *RdRp*-like sequences on the same scaffold no further than 1.5 Kbp apart (Table 1; Supplemental Figure 2). IRESPred (Kolekar et al., 2016) was utilized to identify internal ribosomal entry sites (IRES) on putative dinoRNAV EVE with whole sequence integrations.

ORFs were predicted and annotated from dinoRNAV EVE-containing scaffolds and all dinoflagellate chromosomes using Prodigal (Hyatt et al. 2010) and MAKER2 annotation pipeline

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.11.487905; this version posted April 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(Holt and Yandell 2011) with the AUGUSTUS gene prediction software (Stanke et al. 2006). Translated ORFs were then aligned to a hybrid database containing the UniProt/Swiss-Prot database and protein version of RVDB (v.19; DIAMOND-BLASTp). ORFs on putative dinoRNAV EVE-containing scaffolds and chromosomes were further annotated using InterProScan (v5.48-83.0, Pfam, PANTHER) to identify sequences proximal to putative dinoRNAV integrations.

Phylogenetic analysis of dinoRNAV EVEs

Amino acid-based phylogenetic trees were generated with dinoRNAV EVE ORFs (*MCP* and *RdRp*) from scaffold-scale genomic assemblies, metagenomes, transcriptomes, and sequences from exogenous and closely related +ssRNA reference viruses (Supplemental Table 1A,B, Supplementary Table 2). Sequences were aligned using the best fit algorithm determined by MAFFT (v7.464; Katoh and Standley 2013) and reviewed and trimmed manually in MEGA (v7; Kumar et al. 2016). Maximum-likelihood trees were generated with IQTREE2 (Minh et al. 2020) using the model determined by ModelFinder (Kaylaanamoorthy et al. 2017) and 50,000 parametric bootstraps (Hoang et al. 2018) with nearest neighbor interchange optimization.

Results and Discussion

Evidence of Endogenized dinoRNAVs in Coral Holobiont Metagenomes

Putative dinoRNAV EVEs were detected in metagenomes generated from 42 cnidarian holobionts out of 269 sampled across the South Pacific Ocean. The majority of endogenized dinoRNAVs were identified in hydrocoral metagenomes (*Millepora* spp.) which predominantly harbored *Symbiodinium* dinoflagellates (n=105; 70.5%), but EVE-like sequences were also observed in scleractinian coral metagenomes (*Pocillopora* spp.) which predominantly harbored *Cladocopium* and *Durusdinium* dinoflagellates (n=15; 29.5%; Figure 1B,C). No dinoRNAV-like sequences were detected among *Porites* spp. metagenomes (Figure 1, Figure 2). Hydrocoral metagenomes were sequenced at equivalent depths as scleractinian corals and had comparable levels of annotation (Supplementary Table 3); thus, higher dinoRNAV EVE prevalence in hydrocoral libraries was likely not a result of methodological bias. Of the 11 evaluated South Pacific islands, dinoRNAV EVEs were identified in samples from eight (Guam, Gambier, Moorea, Cook, Niue, Malpelo, Coïba, and Las Perlas), spanning 18 unique sites (Figure 1C). There was a distinct longitudinal trend among *Pocillopora* spp. metagenomes; putative dinoRNAV EVEs were only identified in this coral genus on the Central American coast (CAMR, Coastal Pacific Longhurst Province).

Importantly, endogenized dinoRNAV open reading frames (ORFs) appeared to be immediately adjacent to ORFs identified as dinoflagellate (typically Symbiodiniaceae) genes—they were not proximal to coral genes or those of other cellular organisms abundant in these metagenomes (Supplemental Table 4). We examined the Symbiodiniaceae ITS2 profiles (Hume et al. 2020) associated with each metagenome and found that putative dinoRNAV EVEs were primarily associated with *Symbiodinium*, *Cladocopium*, and *Durusdinium,* which exhibited variation on both host and regional scales (Figure 1D). DinoRNAV EVEs were more common in *Symbiodinium*-dominated cnidarians ($F_{2,1044}=25.8$, $p<0.0001$, nested ANOVA; Supplemental Figure 3) relative to cnidarians hosting other Symbiodiniaceae genera, regardless of host. This suggested that dinoRNAV integration may be particularly recurrent or conserved within the genus *Symbiodinium* (Figure 1).

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.11.487905; this version posted April 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

To determine if these putative viral integrations were specific to cnidarian holobiont metagenomes and ensure that they were not artifacts of shared sample processing and sequencing procedures of the Tara Pacific pipeline, we also analyzed seawater metagenomes and publicly available metagenomes from the stony coral-dinoflagellate holobiont, *Acropora* spp. (Supplemental Table 1B). Examination of 120 Tara Oceans pelagic seawater metagenomes (Pesant et al. 2015) yielded no sequences sharing homology to dinoRNAVs. The concentration of Symbiodiniaceae cells within cnidarian tissues is significantly higher than that of the surrounding seawater (Littman et al. 2008, Schuefen et al. 2017, Fujise et al. 2021, Grupstra et al. 2021). Thus, lack of detection of dinoRNAV-like sequences from seawater metagenomes is likely due to reduced genomic signal of Symbiodiniaceae in the water column. Analysis of the 30 non-Tara *Acropora* holobiont metagenomes identified 29 more putative dinoRNAV EVEs (Figure 2). These dinoRNAV EVEs were again neighboring dinoflagellate ORFs. While the Caribbean *Acropora* metagenomes analyzed contained too few reads to resolve the dominant Symbiodiniaceae present, earlier studies of the same coral colonies identified *Symbiodinium* spp. as the primary symbiont present (Muller et al. 2018).

The identification of endogenized dinoRNAV-like sequences in cnidarian holobiont metagenomes, combined with the proximity of dinoRNAV-like ORFs to dinoflagellate-like sequences across metagenomes harboring diverse dinoflagellate consortia, collectively indicate that dinoRNAV EVEs are widespread among Symbiodiniaceae genera (Figure 2 cyan dots).

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.11.487905; this version posted April 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

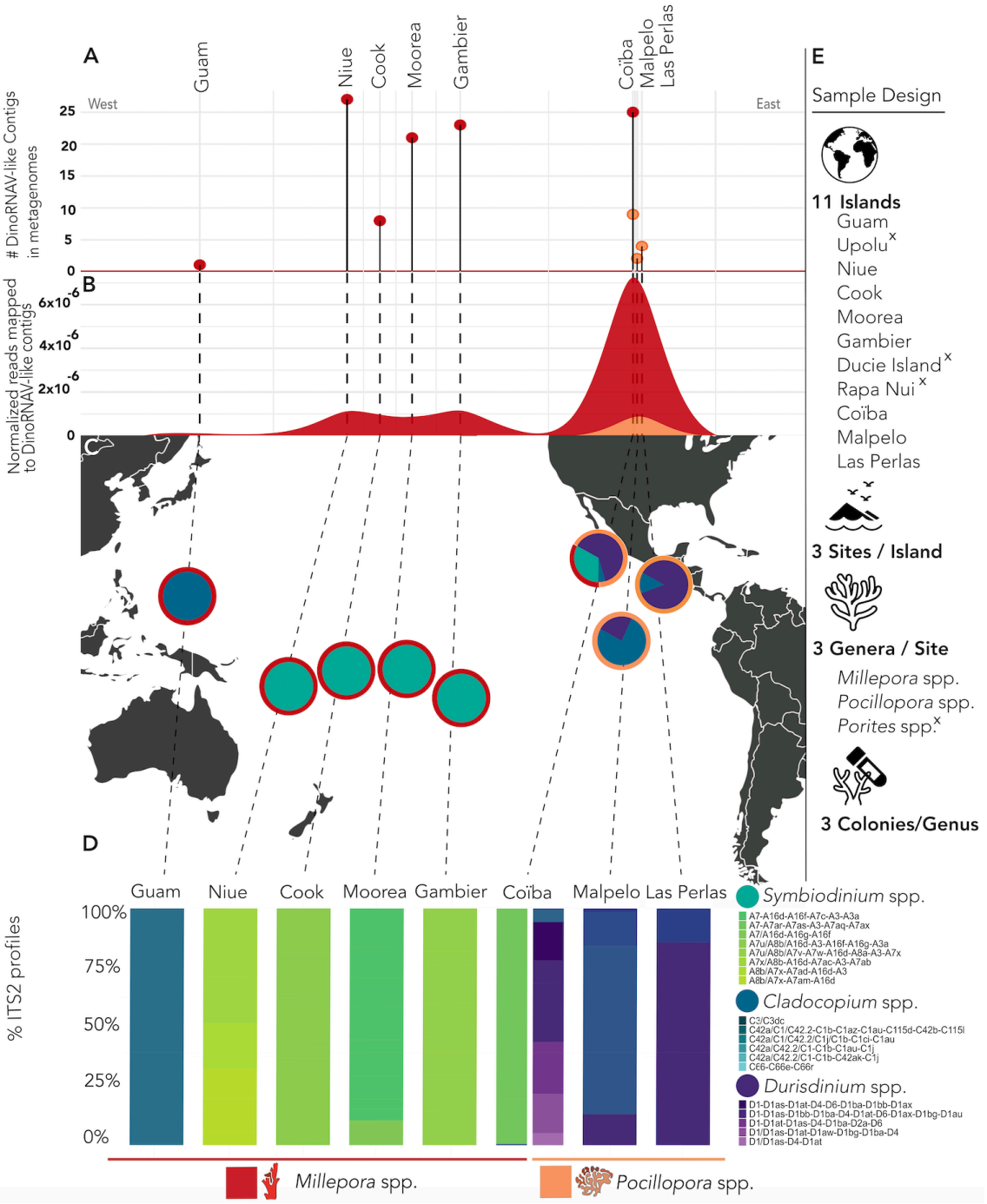
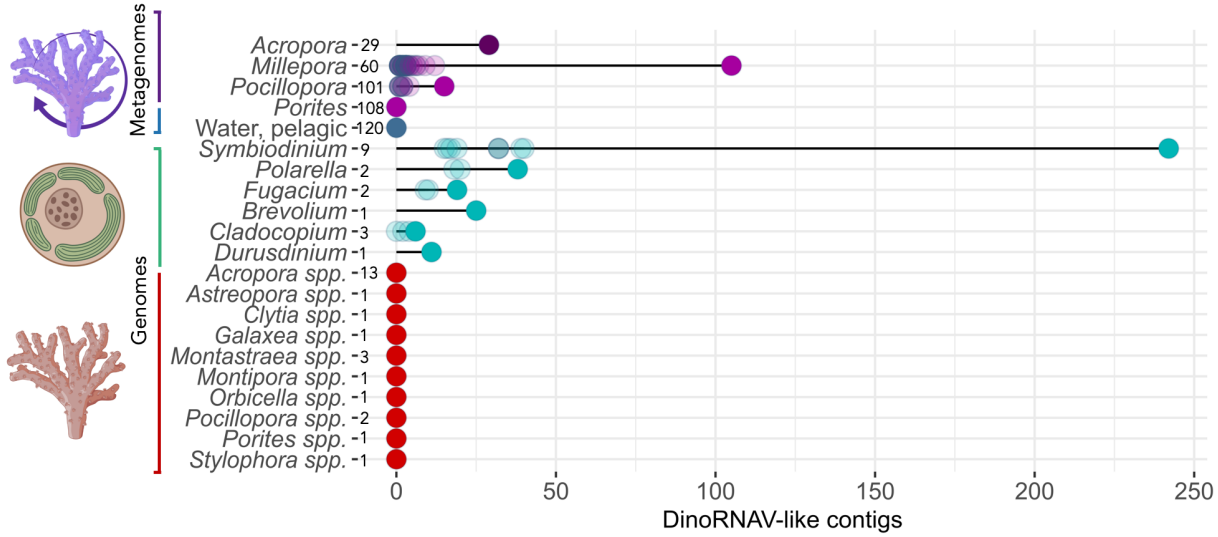


Figure 1. Islands and species (cnidarian and dinoflagellate) correlating with dinoRNAV EVE-like sequence detection among Tara Pacific metagenomes. (A) Count of scaffolds with putative endogenized dinoRNAV-like sequences among Tara Pacific metagenomes, grouped by island and spaced longitudinally by location sampled. (B) Reads mapped to dinoRNAV-like scaffolds within individual Tara Pacific metagenomic libraries, normalized by quality-controlled reads. (C) Sampling sites of Tara Pacific metagenomes explored for endogenized dinoRNAV-like sequences in this study. Internal circles indicate dominant Symbiodiniaceae genera based on ITS2 type profiles, outer ring denotes coral host(s) sampled at each island. (D) Symbiodiniaceae ITS2 type profile metabarcoding as delineated via Symportal (Hume et al, 2019) within island and host. (E) Sample design of Tara Pacific libraries queried for dinoRNAV EVEs. [x] indicates islands or species where no dinoRNAV-like sequences were detected. *[this image is a low resolution placeholder]*

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.11.487905; this version posted April 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Endogenized DinoRNAVs Detected in Symbiodiniaceae Genomes

To further test the hypothesis that dinoRNAVs on reefs infect dinoflagellate symbionts and not cnidarians, we examined 18 scaffold-scale genome assemblies representing the dinoflagellate families Symbiodiniaceae and Suessiaceae as well as 25 cnidarian genomes spanning 10 genera (Supplemental Table 1B; Figure 2; Table 1). Alignments revealed no evidence of endogenized dinoRNAVs in any of the 151,782 aposymbiotic (dinoflagellate-free) cnidarian scaffolds. In contrast, the same approach uncovered 351 (of 593,433) dinoflagellate scaffolds with evidence of endogenized dinoRNAVs (Figure 2; Table 1). The identified 351 dinoRNAV EVE-containing scaffolds were observed across 17 of the 18 dinoflagellate genome assemblies (Table 1). DinoRNAV EVEs were also observed in two assemblies from the free-living dinoflagellate genus, *Polarella* (family Suessiaceae), which is closely related to the family Symbiodiniaceae, and served as an outgroup in this study (Janouškovec et al. 2017; Stephens et al. 2020). Interestingly, assemblies belonging to *Symbiodinium*, the most ancestral Symbiodiniaceae genus (LaJeunesse et al. 2018), contained a higher number of scaffolds with putative dinoRNAV EVEs ($\bar{x}$=28.11, stdev=10.7) relative to assemblies of other Symbiodiniaceae genera ($\bar{x}$=8.71, stdev=11; Figure 2 cyan dots; Table 1). This result may clarify why observations of dinoRNAV-like ORFs were more common in metagenomes dominated by *Symbiodinium* (Figure 1D). The dinoflagellate genome assembly with no detected dinoRNAV EVEs belonged to a relatively incomplete assembly of *Cladocopium* C15, which had the second lowest N50 and lowest BUSCO completeness score of all genomes examined (completeness 11.6%, relative to the average 24.54%; Table 1, Supplementary Table 5). The lower coverage/completeness of the *Cladocopium* C15 assembly indicates a reduced window into this genome. It is therefore possible that when a more complete assembly is generated, dinoRNAV EVE-like sequences will be detectable from this dinoflagellate. However, a linear model suggested that there was no relationship between dinoRNAV EVE detection and assembly statistics (i.e. number of scaffolds, N50, completeness). Furthermore, since we were unable to detect dinoRNAV EVEs in *Porites* metagenomes – a coral species primarily harboring *Cladocopium* C15 symbionts – we hypothesize that dinoRNAV endogenization is less common in this lineage of Symbiodiniaceae.



bioRxiv preprint doi: https://doi.org/10.1101/2022.04.11.487905; this version posted April 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 2. Total quantity of putative endogenized dinoRNAV EVEs identified, broadly organized by sample source (metagenome or genome), and number of libraries or assemblies queried (numbers follow a dash to the right of source name). Opaque circles denote the sum total of dinoRNAV EVE-like sequences identified from each source, while transparent circles denote independent counts per library queried.

	Dinoflagellate Species (strain)	Total # Scaffolds	Host	Location	BUSCO score	dinoRNAV EVE ORFs on scaffolds		
						RdRp	*MCP*	Both
Symbiodiniaceae	*Symbiodinium linucheae* (CCMP2456) [1]	37,772	*Plexaura homamalla*	Bermuda	21.8%	39	0	1
	Symbiodinium microadriaticum (04-503SCI.03) [1]	57,558	*Orbicella faveolata*	Florida, USA	41.6%	30	1	3
	Symbiodinium microadriaticum (CassKB8) [1]	67,937	*Cassiopea sp.*	Hawaii, USA	73.3%	29	1	3
	Symbiodinium microadriticum (CCMP2467) [2,3]	9,688	*Stylophora pistillata*	Red Sea	15.6%	29	1	3
	Symbiodinium natans (CCMP2548) [1]**	2,855	N/A (Isolated from seawater)	Hawaii, USA	15.5%	14	1	3
	Symbiodinium necroappetens (CCMP2469) [1]*	104,583	*Condylactis gigantea*	Jamaica	22.8%	37	4	2
	Symbiodinium pilosum (CCMP2461) [1]**	48,302	*Zoanthus sociatus*	Jamaica	19.8%	15	0	0
	Symbiodinium tridacnidorum (CCMP2592) [1]	6,245	*Heliofungia actiniformis*	Australia	21.1%	17	1	0
	Symbiodinium tridacnidorum (sh18 A3 Y106) [4]	16,176	*Tridacna crocea*	Japan	19.8%	20	1	0
	Brevolium minutum (Mf1.05b) [5]	21,899	*Orbicella faveolata*	Florida, USA	14.2%	21	3	1
	Cladocopium C15 [6]	34,589	*Porites lutea*	Australia	11.6%	0	0	0
	Cladocopium goreaui [7]	41,289	*Acropora tenuis*	Australia	27.7%	4	0	0
	Cladocopium sp C92 (Y103) [4]	6,686	*Fragum* sp.	Japan	19.5%	2	0	0
	Durusdinium trenchii [9]	19,593	*Favia speciosa*	Japan	28.7%	10	1	0
	Fugacium kawagutti (CS156 CCMP2468) [7]	16,959	*Montipora verrucosa*	Hawaii, USA	8.3%	8	1	0
	Fugacium kawagutti (CCMP2468) [9]	30,040	*Montipora verrucosa*	Hawaii, USA	17.9%	9	1	0
	n=314 scaffolds with DinoRNAV EVE-like sequences							
Suessiaceae	*Polarella glacialis* (CCMP1383) [10] **	33,494	N/A (Free-living, isolated from seawater)	Antarctica	20.8%	20	0	0
	Polarella glacialis (CCMP2088) [10] **	37,768	N/A (Free-living, isolated from seawater)	Arctic	21.8%	18	0	0
	n=38 total scaffolds with DinoRNAV EVE-like sequences							

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.11.487905; this version posted April 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 1. DinoRNAV EVE-like detections from representative Symbiodiniaceae and Suessiaceae dinoflagellate scaffold-level genome assemblies, as well as the host species and location of isolation for each dinoflagellate. Assembly coverage and completeness are measured via BUSCO score (% completeness, or %C; Simão et al, 2015). * Indicates species with documented opportunistic life history; ** Indicates species with documented free-living life history per principal species description. Total counts of dinoflagellate scaffolds in genomes queried with individual endogenized dinoRNAV ORFs (*RdRp, MCP*) or both ORFs nearby one another are provided. *RdRp* = RNA-dependent RNA polymerase; *MCP* = major capsid protein. [1] Gonzalez-Pech et al. 2021, [2] Aranda et al. 2016, [3] Gonzalez-Pech et al. 2019, [4] Shoguchi et al. 2018, [5] Shoguchi et al. 2013, [6] Robbins et al. 2019, [7] Liu et al. 2018, [8] Shoguchi et al. 2020, [9] Lin et al 2015, [10]Stephens et al. 2020. Further genome citations (including accession numbers) and BUSCO completion metrics can be found in Supplementary Table 5.

Incomplete ORFs and Possible Duplications Indicate Endogenization of DinoRNAVs

The repeated observation of putative dinoRNAV EVEs in dinoflagellate scaffolds and contigs from metagenomes and genomes suggests these sequences are either (1) conserved sequence artifacts of Symbiodiniaceae-dinoRNAV interactions, and/or (2) evidence of highly prevalent dinoflagellate viruses, commonly integrated and propagated via their single-celled hosts. If the observed dinoRNAV-like sequences represent active infections capable of generating virions during egress, we would, at minimum, expect essential ORFs associated with replication (RNA-dependent RNA polymerase, *RdRp*) and virion structure (Major Capsid Protein, *MCP*) to be endogenized on the same scaffold. We would additionally expect to observe overall conservation of ORF length/composition (with a lack of internal stop codons or significant deletions) when aligning the dinoRNAV-like sequences detected here with known exogenous dinoRNAV sequences.

However, both DIAMOND and gene prediction analyses generally depicted dinoRNAV-like ORFs in isolation on separate scaffolds. While 28 MCP and 73 RdRp dinoRNAV ORFs were annotated, both ORFs were present on a Symbiodiniaceae scaffold – potentially representing whole dinoRNAV genome integrations – in only 14 instances. Thirteen of these 14 were from *Symbiodinium* genomes, whereas one scaffold was from *Breviolum minutum*, a member of the second most ancestral dinoflagellate genus (Table 1; LaJeunesse et al. 2018). To assess the conservation of putative dinoRNAV EVE sequence length/composition, we aligned the genomic and single ORF EVEs to reference exogenous dinoRNAV sequences. The reference genome for reef-associated dinoRNAVs is ~5 Kbp long and contains a 1,071 bp noncoding region between ORFs, with a 124-nucleotide internal ribosomal binding site (Levin et al. 2017). In this study, for 13 of the scaffolds in which dinoRNAV ORFs were detected, the putative noncoding region between the *MCP* and *RdRp* EVEs ranged from ~200-800 bp (except for a scaffold belonging to *S. linucheae* CCMP2456, which contained a ~79 kbp noncoding region, and was excluded in further alignments). No internal ribosomal binding sites were detected within the putative dinoRNAV EVEs identified in dinoflagellate genomes. A nucleotide-based alignment to Levin et al.'s (2017) reference dinoRNAV genome indicated that the putative dinoRNAV EVEs presented here contained substantial insertions and/or deletions (Supplemental Figure 2). Translated exogenous dinoRNAV *MCP* ORFs are reported to be ~358 aa in length (Levin et al. 2017; Figure 3 top sequences), but dinoRNAV-like *MCP* sequences recovered in this study ranged from 116-605aa in length. Furthermore, comparisons of these endogenous *MCPs* to exogenous reference sequences revealed internal stop codons and overall low similarity (Figure 3), instead sharing high conservation in structural amino acid motifs. Amino acid-based alignment of endogenous dinoRNAV MCPs to metatranscriptome- and amplicon-generated exogenous reference sequences (Levin et al. 2017, Montalvo-Proaño et al. 2017) revealed that

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.11.487905; this version posted April 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

indels and regions of low similarity were observed between three conserved regions across both endogenous and exogenous MCP sequences (red boxes in Figure 3).

Interestingly, multiple whole dinoRNAV integrations were sometimes observed in a single dinoflagellate genome. For example, genome assemblies of four different *S. microadriacticum* strains contained two or three whole dinoRNAV EVEs each (Table 1; Figure 3).

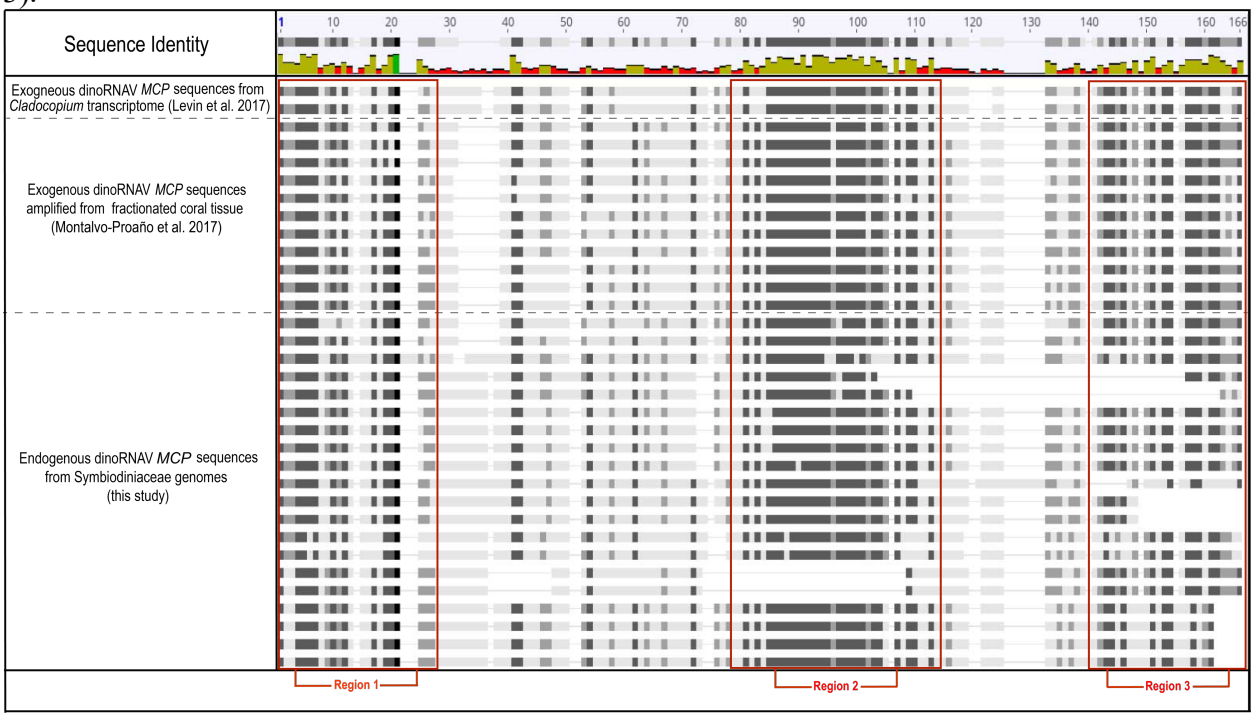


Figure 3. Amino acid alignment including major capsid protein (*MCP*) reference sequences from exogenous dinoRNAV-like +ssRNA viruses, as well as putatively endogenous dinoRNAV *MCP* sequences. Exogenous reference sequences include: 1. Symbiodiniaceae +ssRNA virus *MCP* ORFs recovered from a *Cladocopium* sp. transcriptome (Levin et al, 2017), and 2. dinoRNAV *MCP* amplicons from fractionated coral tissue (Montalvo-Proaño et al. 2017). Conserved regions observed in exogenous and putatively endogenous sequences are labeled as Regions 1-3.

Pairwise alignments measuring shared nucleotide identity of whole dinoRNAV EVEs across Symbiodiniaceae scaffolds revealed that the *S. microadriaticum* genomes and the *S. necroappetens* genome share two whole genome dinoRNAV EVEs (provisionally dinoRNAV-A and dinoRNAV-B; Supplemental Figure 2; Clustal-Omega; Sievers et al. 2011). *S. microadriaticum* dinoRNAV-B was identical in all strains and shared 97% identity with the *S. necroappetens* dinoRNAV-B, yet proximal genes varied (Supplemental Table 6). The inconsistent composition and fragmented nature of both the genomic and single ORF dinoRNAV EVEs reported here supports the hypothesis that these sequences are not capable of generating replicative virions and are best interpreted as multiple integrations of dinoRNAVs into a host genome.

A Potential Mechanism for dinoRNAV Endogenization: Host-Provisioned Retroelements

To assess if general genomic "neighborhoods" are conserved across dinoRNAV integrations (e.g. site location and synteny) and to better understand the genes proximal to EVEs on Symbiodiniaceae genomes, a chromosome-scale *Symbiodinium microadriaticum* genome

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.11.487905; this version posted April 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

assembly was evaluated (Figure 4). The highest quality dinoflagellate genome assembly currently available revealed dinoRNAV-like ORFs on 18 of 94 chromosomes, with at least one *RdRp* on each, and some with multiple (two with n=2 *RdRps*, three with n=3 *RdRps*). On three of the chromosomes (# 30, 35, and 74), there were predicted ORFs annotated as dinoRNAV *MCPs* in close proximity to a *RdRp* ORF (separated by noncoding regions 319-656nt), indicative of a potential full-length dinoRNAV genome integration. These results corroborate detections of multiple genomic dinoRNAV EVEs in scaffold-scale assemblies of *Symbiodinium microadriticum* genomes (Supplemental Figure 2). The higher-resolution *S. microadriaticum* chromosome-level assembly facilitated the identification of an additional dinoRNAV genomic EVE (n=4 for chromosome-level vs. n=3 for scaffold-level, Supplemental Figure 2), two of which were identified on Chromosome 74 and were separated by 2,501 nucleotides. Of note, Nand et al. (2021) reported a decreasing abundance and expression of genes towards the center of chromosomes (past ~2Mpb of a telomere), where there was an increase in repetitive elements; this is where 26 of 29 putative dinoRNAV EVEs were identified in the chromosome-level assembly. Furthermore, ORFs neighboring integrations often varied widely both in proximity and predicted function: from collagen and RNA binding protein to reverse transcriptase and non-LTR retrotransposable elements, these ORFs potentially contributed to the integration of the putative dinoRNAV EVEs.

Host-provisioned retroelements are a proposed mechanism of non-retroviral RNA virus endogenization (Horie et al, 2010, Flynn & Moreau 2019). We examined the annotated ORFs surrounding dinoRNAV-like sequences on the 18 EVE-containing *S. microadriaticum* chromosomes to gain insight into the integration of dinoRNAVs. *Symbiodinium* contain numerous long interspersed nuclear elements (LINEs) relative to other Symbiodiniaceae genera, with LINES comprising 74.10-171.31 Mbp of *Symbiodinium* genomes, relative to an average of 7.48 Mbp of the genomes of in other genera (González-Pech et al. 2021, Nand et al. 2021, Mita and Boeke 2016). This group of non-long terminal repeat (non-LTR) eukaryotic retrotransposons contains reverse transcriptases and is conserved, implying that they are active and may facilitate non-retroviral endogenization of dinoRNAVs within Symbiodiniaceae. Proximal to putative dinoRNAV *MCP* and *RdRp* ORFs on *S. microadriaticum* chromosomes, ~40% of annotated ORFs (35 of 88 annotated proteins) were similar to non-LTR retrotransposable elements seen in other eukaryotic genomes (Figure 4, Supplemental table 6), sometimes <300bp 5' upstream. We also annotated ORFs similar to hypothetical virus proteins, suggesting that this mechanism may facilitate integration of sequences beyond dinoRNAVs. These observations of non-LTR retrotransposable elements, sometimes in very close proximity to dinoRNAV EVEs, supports the hypothesis of cis-acting host-driven integration and may explains the increased abundance of dinoRNAV EVEs, particularly in *Symbiodinium* genomes.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.11.487905; this version posted April 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

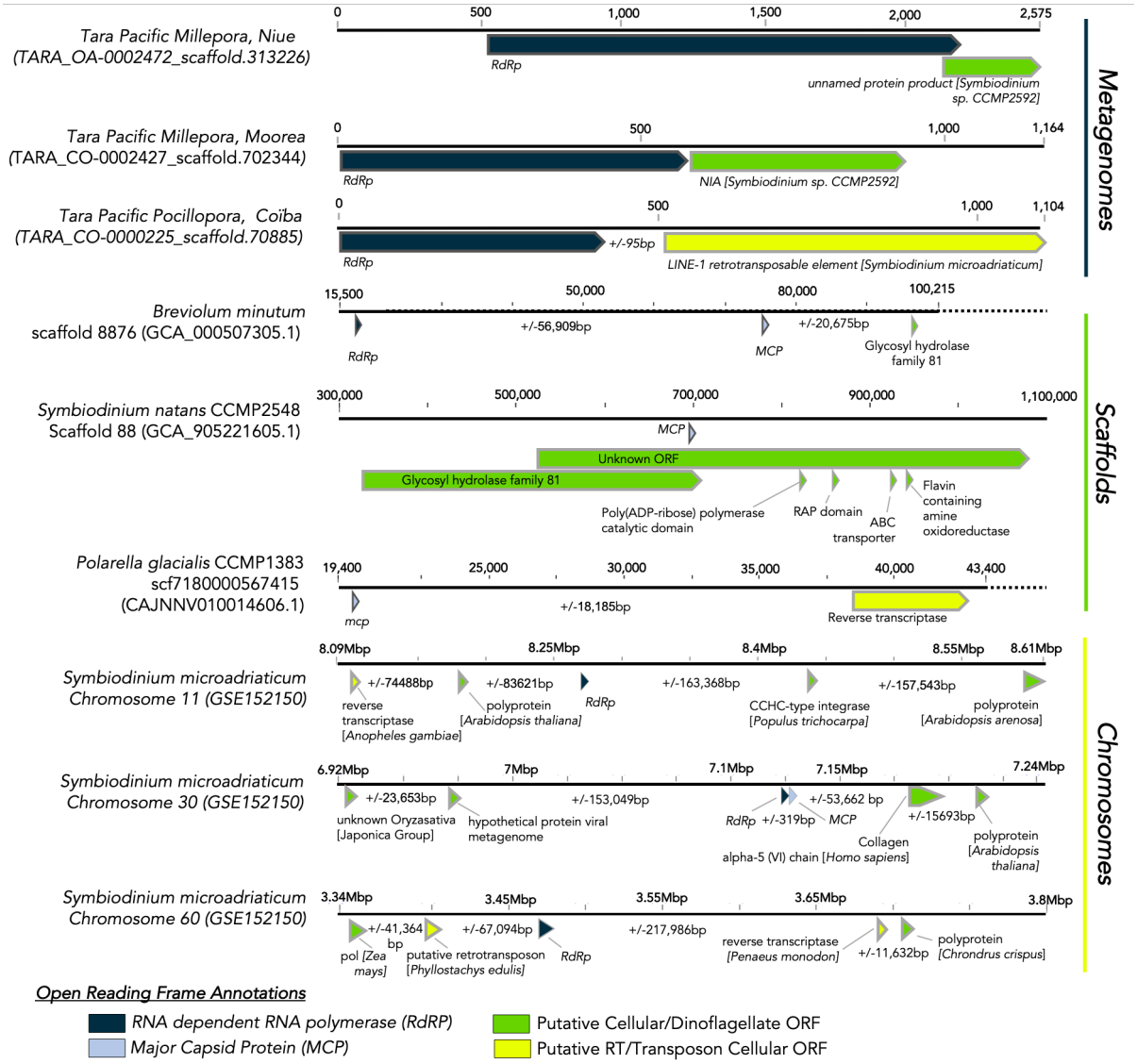


Figure 4. Representative scaffolds and chromosome fragments containing putative dinoRNAV EVEs (*MCP*, light blue; *RdRp*, navy blue ORFs with complete description in supplemental table 6). Open reading frame (ORF) color broadly indicates cellular versus putative +ssRNA viral homology; yellow ORFs may be exploited mechanisms for viral integration. (+/-) base pair values represent sequence lengths between ORFs.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.11.487905; this version posted April 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

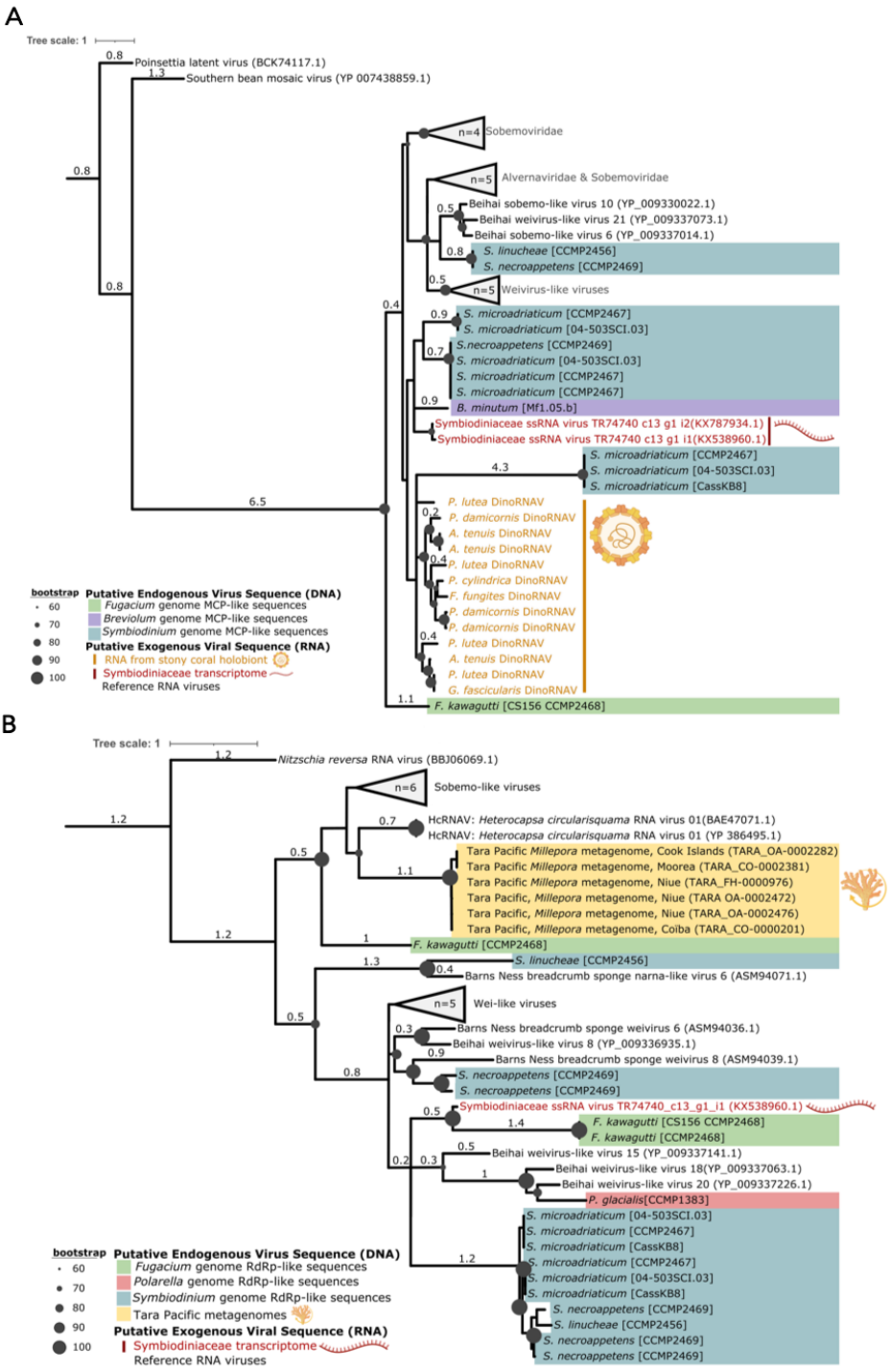


Figure 5. Phylogenies of dinoRNAV Major Capsid Protein (*MCP,* **A**) and RNA-dependent RNA polymerase (*RdRp,* **B**) ORFs recovered from host metagenomes, transcriptomes, polished genomes, and extant +ssRNA reference viruses from amplicon libraries (A only). Both trees include *Dinornavirus* reference sequences. (A) Maximum-likelihood tree of *MCP* amino acid sequences generated with a LG+F+G4 substitution model and 50,000 parametric bootstraps, illustrating the similarity of putative dinoRNAV EVEs (this study) to extant dinoRNAVs from stony coral colonies. (B) Maximum-likelihood tree of *RdRp* amino acid sequences generated with a Blosum62+G4 substitution model and 50,000 parametric bootstraps, demonstrating the similarity of metagenomic dinoRNAV EVE *RdRps* to *RdRps* of the sole recognized *Dinornavirus*, Heterocapsa circularisquama RNA virus (HcRNAV), as well as alignment to each other. Trees were visualized in iTOL. *[this image is a low resolution placeholder]*

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.11.487905; this version posted April 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

DinoRNAV EVEs Show Homology to Extant Exogenous Viruses

Endogenized viruses can lend insight into the dynamics between cellular hosts and extant, free virions. Because many EVEs evolve at the rate of the host genome, rather than at the much faster rate of exogenous +ssRNA viral genomes, EVEs can serve as a snapshot of viral ancestry (Holmes et al. 2010). We compared translated putative dinoRNAV EVEs from this study to putative exogenous dinoRNAVs and extant *Dinornavirus* taxa to better understand if the former were highly conserved sequences (with potential utility to the host) and/or recently integrated EVEs. We found that amino acid translations of endogenous dinoRNAV *MCP* sequences contained conserved motifs observed in the exogenous *MCP* sequences (e.g. regions 1-3 in Figure 3), yet the associated phylogeny was highly polyphyletic along inferred ancestral nodes (Figure 5A, B). Several putative dinoRNAV EVEs shared similarity to extant *MCP*s identified from unfractionated stony coral holobionts via amplicon sequencing (Montalvo-Proaño et al. 2017); these sequences formed an independent, disorganized clade (Figure 5A clade containing yellow and blue sequences), relative to those recovered from dinoflagellate genomes or other invertebrate hosts. *MCP* and *RdRp* ORFs putatively derived from the same dinoflagellate genomes often shared clades (clades containing multiple blue or green sequences in Figure 5A, B), perhaps indicative of duplications within genomes or multiple integration events of particular dinoRNAV lineages within host genera. Symbiodiniaceae first diversified from the psychrophilic, free-living *Polarella* outgroup ~160 million years ago (Janouškovec et al. 2017, Stephens et al. 2020; LaJeunesse et al. 2018). The detection of putative dinoRNAV *RdRp* ORFs within *Polarella* genomes is therefore indicative of either the antiquity of dinoRNAV-dinoflagellate interactions and/or a propensity for dinoRNAV integration across Dinophyceae families. However, the exclusion of the *P. glacialis* dinoRNAV-*RdRp* from *RdRps* of other dinoflagellate clades (pink, Figure 5B) further illustrates the congruence between EVEs and their host genomes.

The expression and functional potential of endogenized dinoRNAVs (if any) remains unclear. Two transcripts derived from *Cladocopium* transcriptomes and annotated as *MCP* ORFs of +ssRNA viral sequences ('TR74740_c13-g1_i1' and 'TR74740_c13-g1_i2', Levin et al. 2017, red text in Figure 5A) shared a clade with putative *Symbiodinium* dinoRNAV EVEs. Likewise, the *RdRp* ORF of 'TR74740_c13-g1_i1' and the *RdRp* of 'GAKY01194223.1'— a transcript derived from a cultured *Symbiodinium microadriaticum* A1 transcriptome—shared similarity to putative endogenous dinoRNAVs (Figure 5B; Levin et al. 2017, Baumgarten et al. 2013). Importantly, both RNA transcripts also shared features characteristic of dinoflagellates, such as a 5' spliced leader sequence ("DinoSL"; Zhang et al. 2013) or dinoflagellate sequence space flanking the dinoRNAV itself (Baumgarten et al. 2013). While 'TR74740_c13-g1_i1' appeared to be in the top 0.03% of expressed transcripts at under certain thermal conditions, GAKY01194223.1 appeared exhibit moderately differential expression at the extremes of temperature and ionic stress in a cultured host (Levin et al. 2017, Baumgarten et al. 2013).

While viral *RdRps* have been leveraged by eukaryotes in multiple pathways (Lipardi and Paterson 2010), the fragmented nature of the putative dinoRNAV EVEs *in silico* may contribute to a role in triggering antiviral mechanisms in their hosts (Blair et al. 2020, Suzuki et al. 2020). Given that the *Symbiodinium* genome contains all core RNAi protein machinery, including Argonaute and Dicer, and that GAKY01194223.1 folds into several hairpins (ΔG = -142.5kcal/mol; Supplemental Figure 4 examples), Symbiodiniaceae may use the putative EVE ncRNA identified here to develop host immunity against extant, exogenous dinoRNAVs.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.11.487905; this version posted April 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Furthermore, *S. microadriaticum* harboring dinoRNAV EVEs contained numerous non-retroviral EVEs of other viral families (Supplemental Figure 5) in close proximity, such as *Herpesviridae, Baculoviridae, Poxviridae, Iridoviridae, Phycodnaviridae, Pandoraviridae* and *Pithoviridae*, ssDNA viruses of the family *Shotokuvirae*, -ssRNA viruses from the family *Rhabdoviridae* and +ssRNA viruses from the family *Coronaviridae* (Supplemental Figure 5). Tara Pacific metagenomes corroborate findings of similar *RdRps* from these viral families (Supplemental Figure 5).

Conclusions

Endogenized viral elements, such as dinoRNAVs, demonstrate how *in silico* identification can provide context for viral genomes in non-model, symbiotic systems such as coral holobionts, impacting how we study coral reefs and their viral consortia. Endogenized viral elements (EVEs) have been effectively utilized in many terrestrial systems to better understand the evolutionary history of viruses ("paleovirology") and pair hosts with extant viruses in multipartite systems. We propose that dinoRNAVs utilize host-provisioned mechanisms (e.g., LINEs) to integrate into single-celled dinoflagellate genomes as EVEs. We detected heritable integrations of multiple putative dinoRNAV genes in Symbiodiniaceae scaffolds from cnidarian metagenomes, as well as in diverse genomes of cultured Symbiodiniaceae; no integrations were detected from seawater metagenomes nor diverse aposymbiotic cnidarian genomes. The apparent pervasive nature of dinoRNAV-like sequences among dinoflagellate genomes (especially the genus *Symbiodinium*) suggests widespread and recurrent/ancestral integration and conservation of these EVEs. The findings presented in this study further validate the dinoRNAV-Symbiodiniaceae virus-host pair, enhancing our understanding of ecologically and economically important cnidarian holobionts and opening the door to examining the role of EVEs in reef health.

Data Availability

Metadata and sequences are accessible in zenodo:
https://zenodo.org/communities/tarapacific?page=1&size=20

Acknowledgements and Contributions

Special thanks to the Tara Ocean Foundation, the R/V Tara crew and the Tara Pacific Expedition Participants (https://doi.org/10.5281/zenodo.3777760). We are keen to thank the commitment of the following institutions for their financial and scientific support that made this unique Tara Pacific Expedition possible: CNRS, PSL, CSM, EPHE, Genoscope, CEA, Inserm, Université Côte d'Azur, ANR, agnès b., UNESCO-IOC, the Veolia Foundation, the Prince Albert II de Monaco Foundation, Région Bretagne, Billerudkorsnas, AmerisourceBergen Company, Lorient Agglomération, Oceans by Disney, L'Oréal, Biotherm, France Collectivités, Fonds Français pour l'Environnement Mondial (FFEM), Etienne Bourgois, and the Tara Ocean Foundation teams. Tara Pacific would not exist without the continuous support of the participating institutes. This research is further supported by NSF OCE #1635798 to AMSC, NSF DOB Grant 2025457 to RLVT, and with additional support from NSF PRFB – 1907184 to KSIB.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.11.487905; this version posted April 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

List of Tara Pacific Consortium Coordinators

Sylvain Agostini (orcid.org/0000-0001-9040-9296) Shimoda Marine Research Center, University of Tsukuba, 5-10-1, Shimoda, Shizuoka, Japan
Denis Allemand (orcid.org/0000-0002-3089-4290) Centre Scientifique de Monaco, 8 Quai Antoine Ier, MC-98000, Principality of Monaco
Bernard Banaigs (orcid.org/0000-0003-3473-4283) PSL Research University: EPHE-UPVD-CNRS, USR 3278 CRIOBE, Université de Perpignan, France
Emilie Boissin (orcid.org/0000-0002-4110-790X) PSL Research University: EPHE-UPVD-CNRS, USR 3278 CRIOBE, Laboratoire d'Excellence CORAIL, Université de Perpignan, 52 Avenue Paul Alduy, 66860 Perpignan Cedex, France
Emmanuel Boss (orcid.org/0000-0002-8334-9595) School of Marine Sciences, University of Maine, Orono, 04469, Maine, USA
Chris Bowler (orcid.org/0000-0003-3835-6187) Institut de Biologie de l'Ecole Normale Supérieure (IBENS), Ecole normale supérieure, CNRS, INSERM, Université PSL, 75005 Paris, France
Colomban de Vargas (orcid.org/0000-0002-6476-6019) Sorbonne Université, CNRS, Station Biologique de Roscoff, AD2M, UMR 7144, ECOMAP 29680 Roscoff, France & Research Federation for the study of Global Ocean Systems Ecology and Evolution, FR2022/ Tara Oceans-GOSEE, 3 rue Michel-Ange, 75016 Paris, France
Eric Douville (orcid.org/0000-0002-6673-1768) Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France
Michel Flores (orcid.org/0000-0003-3609-286X) Weizmann Institute of Science, Department of Earth and Planetary Sciences, 76100 Rehovot, Israel
Didier Forcioli (orcid.org/0000-0002-5505-0932) Université Côte d'Azur, CNRS, INSERM, IRCAN, Medical School, Nice, France and Department of Medical Genetics, CHU of Nice, France
Paola Furla (orcid.org/0000-0001-9899-942X) Université Côte d'Azur, CNRS, INSERM, IRCAN, Medical School, Nice, France and Department of Medical Genetics, CHU of Nice, France
Pierre Galand (orcid.org/0000-0002-2238-3247) Sorbonne Université, CNRS, Laboratoire d'Ecogéochimie des Environnements Benthiques (LECOB), Observatoire Océanologique de Banyuls, 66650 Banyuls sur mer, France
Eric Gilson (orcid.org/0000-0001-5738-6723) Université Côte d'Azur, CNRS, Inserm, IRCAN, France
Fabien Lombard (orcid.org/0000-0002-8626-8782) Sorbonne Université, Institut de la Mer de Villefranche sur mer, Laboratoire d'Océanographie de Villefranche, F-06230 Villefranche-sur-Mer, France
Stéphane Pesant (orcid.org/0000-0002-4936-5209) European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK
Serge Planes (orcid.org/0000-0002-5689-5371) PSL Research University: EPHE-UPVD-CNRS, USR 3278 CRIOBE, Laboratoire d'Excellence CORAIL, Université de Perpignan, 52 Avenue Paul Alduy, 66860 Perpignan Cedex, France

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.11.487905; this version posted April 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

612 Stéphanie Reynaud (orcid.org/0000-0001-9975-6075) Centre Scientifique de Monaco, 8 Quai
613 Antoine Ier, MC-98000, Principality of Monaco
614 Matthew B. Sullivan (orcid.org/0000-0003-4040-9831) The Ohio State University, Departments
615 of Microbiology and Civil, Environmental and Geodetic Engineering, Columbus, Ohio,
616 43210 USA
617 Shinichi Sunagawa (orcid.org/0000-0003-3065-0314) Department of Biology, Institute of
618 Microbiology and Swiss Institute of Bioinformatics, Vladimir-Prelog-Weg 4, ETH
619 Zürich, CH-8093 Zürich, Switzerland
620 Olivier Thomas (orcid.org/0000-0002-5708-1409) Marine Biodiscovery Laboratory, School of
621 Chemistry and Ryan Institute, National University of Ireland, Galway, Ireland
622 Romain Troublé (ORCID not-available) Fondation Tara Océan, Base Tara, 8 rue de Prague, 75
623 012 Paris, France
624 Rebecca Vega Thurber (orcid.org/0000-0003-3516-2061) Oregon State University, Department
625 of Microbiology, 220 Nash Hall, 97331Corvallis OR USA
626 Christian R. Voolstra (orcid.org/0000-0003-4555-3795) Department of Biology, University of
627 Konstanz, 78457 Konstanz, Germany
628 Patrick Wincker (orcid.org/0000-0001-7562-3454) Génomique Métabolique, Genoscope, Institut
629 François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, 91057 Evry, France
630 Didier Zoccola (orcid.org/0000-0002-1524-8098) Centre Scientifique de Monaco, 8 Quai
631 Antoine Ier, MC-98000, Principality of Monaco
632

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.11.487905; this version posted April 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

Aiewsakun, P., & Katzourakis, A. (2015). Endogenous viruses: Connecting recent and ancient viral evolution. *Virology*, *479–480*, 26–37. https://doi.org/10.1016/j.virol.2015.02.011

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Aranda, M., Li, Y., Liew, Y. J., Baumgarten, S., Simakov, O., Wilson, M. C., Piel, J., Ashoor, H., Bougouffa, S., Bajic, V. B., Ryu, T., Ravasi, T., Bayer, T., Micklem, G., Kim, H., Bhak, J., LaJeunesse, T. C., & Voolstra, C. R. (2016). Genomes of coral dinoflagellate symbionts highlight evolutionary adaptations conducive to a symbiotic lifestyle. *Scientific Reports*, *6*(1), 39734. https://doi.org/10.1038/srep39734

Ballinger, M. J., Bruenn, J. A., & Taylor, D. J. (2012). Phylogeny, integration and expression of sigma virus-like genes in *Drosophila*. *Molecular Phylogenetics and Evolution*, *65*(1), 251–258. https://doi.org/10.1016/j.ympev.2012.06.008

Baumgarten, S., Bayer, T., Aranda, M., Liew, Y. J., Carr, A., Micklem, G., & Voolstra, C. R. (2013). Integrating microRNA and mRNA expression profiling in *Symbiodinium microadriaticum*, a dinoflagellate symbiont of reef-building corals. *BMC Genomics*, *14*(1), 704. https://doi.org/10.1186/1471-2164-14-704

Bigot, T., Temmam, S., Pérot, P., & Eloit, M. (2020). RVDB-prot, a reference viral protein database and its HMM profiles. *F1000 Research*, *8*, 530. https://doi.org/10.12688/f1000research.18776.2

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.11.487905; this version posted April 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

654 Blair, C. D., Olson, K. E., & Bonizzoni, M. (2020). The Widespread occurrence and potential
655 biological roles of endogenous viral elements in insect genomes. *Current Issues in Molecular*
656 *Biology*, 13–30. https://doi.org/10.21775/cimb.034.013
657 Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using
658 DIAMOND. *Nature Methods*, *12*(1), 59–60. https://doi.org/10.1038/nmeth.3176
659 Bushnell, B., Rood, J., & Singer, E. (2017). BBMerge – Accurate paired shotgun read merging via
660 overlap. *PLOS ONE*, *12*(10), e0185056. https://doi.org/10.1371/journal.pone.0185056
661 Chiba, S., Kondo, H., Tani, A., Saisho, D., Sakamoto, W., Kanematsu, S., & Suzuki, N. (2011).
662 Widespread endogenization of genome sequences of non-retroviral RNA viruses into plant
663 genomes. *PLOS Pathogens*, *7*(7), e1002146. https://doi.org/10.1371/journal.ppat.1002146
664 Chu, H., Jo, Y., & Cho, W. K. (2014). Evolution of endogenous non-retroviral genes integrated into
665 plant genomes. *Current Plant Biology*, *1*, 55–59. https://doi.org/10.1016/j.cpb.2014.07.002
666 Correa, A. M. S., Ainsworth, T. D., Rosales, S. M., Thurber, A. R., Butler, C. R., & Vega Thurber,
667 R. L. (2016). Viral outbreak in corals associated with an *in situ* bleaching event: atypical herpes-
668 like viruses and a new Megavirus infecting *Symbiodinium*. *Frontiers in Microbiology*, *7*.
669 https://doi.org/10.3389/fmicb.2016.00127
670 Correa, A. M. S., Howard-Varona, C., Coy, S. R., Buchan, A., Sullivan, M. B., & Weitz, J. S.
671 (2021). Revisiting the rules of life for viruses of microorganisms. *Nature Reviews Microbiology*,
672 *19*(8), 501–513. https://doi.org/10.1038/s41579-021-00530-x
673 Correa, A. M. S., Welsh, R. M., & Vega Thurber, R. L. (2013). Unique nucleocytoplasmic dsDNA
674 and +ssRNA viruses are associated with the dinoflagellate endosymbionts of corals. *The ISME*
675 *Journal*, *7*(1), 13–27. https://doi.org/10.1038/ismej.2012.75

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.11.487905; this version posted April 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

676 Feschotte, C., & Gilbert, C. (2012). Endogenous viruses: insights into viral evolution and impact on
677 host biology. *Nature Reviews Genetics*, *13*(4), 283–296. https://doi.org/10.1038/nrg3199
678 Flynn, P. J., & Moreau, C. S. (2019). Assessing the diversity of endogenous viruses throughout ant
679 genomes. *Frontiers in Microbiology*, *10*, 1139. https://doi.org/10.3389/fmicb.2019.01139
680 Frank, J. A., & Feschotte, C. (2017). Co-option of endogenous viral sequences for host cell
681 function. *Current Opinion in Virology*, *25*, 81–89. https://doi.org/10.1016/j.coviro.2017.07.021
682 Fujise, L., Suggett, D. J., Stat, M., Kahlke, T., Bunce, M., Gardner, S. G., Goyen, S., Woodcock,
683 S., Ralph, P. J., Seymour, J. R., Siboni, N., & Nitschke, M. R. (2021). Unlocking the
684 phylogenetic diversity, primary habitats, and abundances of free-living Symbiodiniaceae on a
685 coral reef. *Molecular Ecology*, *30*(1), 343–360. https://doi.org/10.1111/mec.15719
686 Gallot-Lavallée, L., & Blanc, G. (2017). A Glimpse of Nucleo-Cytoplasmic Large DNA Virus
687 biodiversity through the eukaryotic genomics window. *Viruses*, *9*(1), 17.
688 https://doi.org/10.3390/v9010017
689 Glynn, P. W. (1996). Coral reef bleaching: facts, hypotheses and implications. *Global Change*
690 *Biology*, *2*(6), 495–509. https://doi.org/10.1111/j.1365-2486.1996.tb00063.x
691 González-Pech, R. A., Chen, Y., Stephens, T. G., Shah, S., Mohamed, A. R., Lagorce, R.,
692 Bhattacharya, D., Ragan, M. A., & Chan, C. X. (2019). *Genomes of Symbiodiniaceae reveal*
693 *extensive sequence divergence but conserved functions at family and genus levels* [Preprint].
694 Genomics. https://doi.org/10.1101/800482
695 González-Pech, R. A., Stephens, T. G., Chen, Y., Mohamed, A. R., Cheng, Y., Shah, S., Dougan,
696 K. E., Fortuin, M. D. A., Lagorce, R., Burt, D. W., Bhattacharya, D., Ragan, M. A., & Chan, C.
697 X. (2021). Comparison of 15 dinoflagellate genomes reveals extensive sequence and structural

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.11.487905; this version posted April 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

698 divergence in family Symbiodiniaceae and genus *Symbiodinium*. *BMC Biology*, *19*(1), 73.
699 https://doi.org/10.1186/s12915-021-00994-6
700 Grupstra, C. G. B., Rabbitt, K. M., Howe-Kerr, L. I., & Correa, A. M. S. (2021). Fish predation on
701 corals promotes the dispersal of coral symbionts. *Animal Microbiome*, *3*(1), 25.
702 https://doi.org/10.1186/s42523-021-00086-4
703 Grupstra, C. G. B., L. I. Howe-Kerr, A. J. Veglia, R. L. Bryant, S. R. Coy, P. L. Blackwelder and
704 A. M. S. Correa (2022). Thermal stress triggers productive viral infection of a key coral reef
705 symbiont. *The ISME Journal*. 10.1038/s41396-022-01194-y
706 Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for
707 genome assemblies. *Bioinformatics*, *29*(8), 1072–1075.
708 https://doi.org/10.1093/bioinformatics/btt086
709 Harrison, E., & Brockhurst, M. A. (2017). Ecological and evolutionary benefits of temperate phage:
710 what does or doesn't kill you makes you stronger. *BioEssays*, *39*(12), 1700112.
711 https://doi.org/10.1002/bies.201700112
712 Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2:
713 improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, *35*(2), 518–
714 522. https://doi.org/10.1093/molbev/msx281
715 Holmes, E. C. (2011). The evolution of endogenous viral elements. *Cell Host & Microbe*, *10*(4),
716 368–377. https://doi.org/10.1016/j.chom.2011.09.002
717 Holt, C., & Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database
718 management tool for second-generation genome projects. *BMC Bioinformatics*, *12*(1), 491.
719 https://doi.org/10.1186/1471-2105-12-491

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.11.487905; this version posted April 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

720 Horie, M., Honda, T., Suzuki, Y., Kobayashi, Y., Daito, T., Oshida, T., Ikuta, K., Jern, P.,
721 Gojobori, T., Coffin, J. M., & Tomonaga, K. (2010). Endogenous non-retroviral RNA virus
722 elements in mammalian genomes. *Nature*, *463*(7277), 84–87.
723 https://doi.org/10.1038/nature08695
724 Hume, B. C. C., Smith, E. G., Ziegler, M., Warrington, H. J. M., Burt, J. A., LaJeunesse, T. C.,
725 Wiedenmann, J., & Voolstra, C. R. (2019). SymPortal: A novel analytical framework and
726 platform for coral algal symbiont next-generation sequencing *ITS2* profiling. *Molecular Ecology*
727 *Resources*, *19*(4), 1063–1080. https://doi.org/10.1111/1755-0998.13004
728 Hume, Benjamin C.C., Poulain, Julie, Pesant, Stéphane, Belser, Caroline, Ruscheweyh, Hans-
729 Joachim, Moulin, Clémentine, Boissin, Emilie, Bourdin, Guillaume, Iwankow, Guillaume,
730 Romac, Sarah, Agostini, Sylvain, Banaigs, Bernard, Boss, Emmanuel, Bowler, Chris, de Vargas,
731 Colomban, Douville, Eric, Flores, J. Michel, Forcioli, Didier, Furla, Paola, … Voolstra, Christian
732 R. (2020). Tara Pacific metabarcoding sequencing (16S, 18S, ITS2) reference & replication
733 tables version 1 (Version 1) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.4073035
734 Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010).
735 Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC*
736 *Bioinformatics*, *11*(1), 119. https://doi.org/10.1186/1471-2105-11-119
737 Janouškovec, J., Gavelis, G. S., Burki, F., Dinh, D., Bachvaroff, T. R., Gornik, S. G., Bright, K. J.,
738 Imanian, B., Strom, S. L., Delwiche, C. F., Waller, R. F., Fensome, R. A., Leander, B. S.,
739 Rohwer, F. L., & Saldarriaga, J. F. (2017). Major transitions in dinoflagellate evolution unveiled
740 by phylotranscriptomics. *Proceedings of the National Academy of Sciences*, *114*(2), E171–E180.
741 https://doi.org/10.1073/pnas.1614842114

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.11.487905; this version posted April 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

742 Jebb, D., Huang, Z., Pippel, M., Hughes, G. M., Lavrichenko, K., Devanna, P., Winkler, S.,
743 Jermiin, L. S., Skirmuntt, E. C., Katzourakis, A., Burkitt-Gray, L., Ray, D. A., Sullivan, K. A.
744 M., Roscito, J. G., Kirilenko, B. M., Dávalos, L. M., Corthals, A. P., Power, M. L., Jones, G., …
745 Teeling, E. C. (2020). Six reference-quality genomes reveal evolution of bat adaptations. *Nature*,
746 *583*(7817), 578–584. https://doi.org/10.1038/s41586-020-2486-3
747 Jern, P., & Coffin, J. M. (2008). Effects of Retroviruses on Host Genome Function. *Annual Review*
748 *of Genetics*, *42*(1), 709–732. https://doi.org/10.1146/annurev.genet.42.110807.091501
749 Johnson, W. E. (2010). Endless Forms Most Viral. *PLoS Genetics*, *6*(11), e1001210.
750 https://doi.org/10.1371/journal.pgen.1001210
751 Johnson, W. E. (2015). Endogenous retroviruses in the genomics era. *Annual Review of Virology*,
752 *2*(1), 135–159. https://doi.org/10.1146/annurev-virology-100114-054945
753 Johnson, W. E. (2019). Origins and evolutionary consequences of ancient endogenous retroviruses.
754 *Nature Reviews Microbiology*, *17*(6), 355–370. https://doi.org/10.1038/s41579-019-0189-2
755 Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017).
756 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6),
757 587–589. https://doi.org/10.1038/nmeth.4285
758 Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7:
759 improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780.
760 https://doi.org/10.1093/molbev/mst010
761 Katzourakis, A. (2013). Paleovirology: inferring viral evolution from host genome sequence data.
762 *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1626), 20120493.
763 https://doi.org/10.1098/rstb.2012.0493

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.11.487905; this version posted April 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

764 Katzourakis, A., & Gifford, R. J. (2010). Endogenous viral elements in animal genomes. *PLOS*
765 *Genetics*, *6*(11), e1001191. https://doi.org/10.1371/journal.pgen.1001191
766 Knowlton, N., & Rohwer, F. (2003). Multispecies Microbial Mutualisms on Coral Reefs: The Host
767 as a Habitat. *The American Naturalist*, *162*(S4), S51–S62. https://doi.org/10.1086/378684
768 Kojima, S., Yoshikawa, K., Ito, J., Nakagawa, S., Parrish, N. F., Horie, M., Kawano, S., &
769 Tomonaga, K. (2021). Virus-like insertions with sequence signatures similar to those of
770 endogenous nonretroviral RNA viruses in the human genome. *Proceedings of the National*
771 *Academy of Sciences*, *118*(5). https://doi.org/10.1073/pnas.2010758118
772 Kolekar, P., Pataskar, A., Kulkarni-Kale, U. et al. (2016). IRESPred: Web Server for Prediction of
773 Cellular and Viral Internal Ribosome Entry Site (IRES). *Sci Rep*, 6:27436.
774 https://doi.org/10.1038/srep27436
775 Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis
776 version 7.0 for bigger datasets. *Molecular Biology and Evolution*, *33*(7), 1870–1874.
777 https://doi.org/10.1093/molbev/msw054
778 LaJeunesse, T. C., Parkinson, J. E., Gabrielson, P. W., Jeong, H. J., Reimer, J. D., Voolstra, C. R.,
779 & Santos, S. R. (2018). Systematic revision of symbiodiniaceae highlights the antiquity and
780 diversity of coral endosymbionts. *Current Biology*, *28*(16), 2570-2580.e6.
781 https://doi.org/10.1016/j.cub.2018.07.008
782 Lawrence, S. A., Davy, J. E., Aeby, G. S., Wilson, W. H., & Davy, S. K. (2014). Quantification of
783 virus-like particles suggests viral infection in corals affected by *Porites* tissue loss. *Coral Reefs*,
784 *33*(3), 687–691. https://doi.org/10.1007/s00338-014-1168-8

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.11.487905; this version posted April 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

785 Levin, R. A., Voolstra, C. R., Weynberg, K. D., & van Oppen, M. J. H. (2017). Evidence for a role
786 of viruses in the thermal sensitivity of coral photosymbionts. *The ISME Journal*, *11*(3), 808–812.
787 https://doi.org/10.1038/ismej.2016.154
788 Lin, S., Cheng, S., Song, B., Zhong, X., Lin, X., Li, W., Li, L., Zhang, Y., Zhang, H., Ji, Z., Cai,
789 M., Zhuang, Y., Shi, X., Lin, L., Wang, L., Wang, Z., Liu, X., Yu, S., Zeng, P., … Morse, D.
790 (2015). The *Symbiodinium kawagutii* genome illuminates dinoflagellate gene expression and
791 coral symbiosis. *Science*, *350*(6261), 691–694. https://doi.org/10.1126/science.aad0408
792 Lipardi, C., & Paterson, B. M. (2009). Identification of an RNA-dependent RNA polymerase in
793 *Drosophila* involved in RNAi and transposon suppression. *Proceedings of the National Academy*
794 *of Sciences*, *106*(37), 15645–15650. https://doi.org/10.1073/pnas.0904984106
795 Littman, R. A., van Oppen, M. J. H., & Willis, B. L. (2008). Methods for sampling free-living
796 *Symbiodinium* (zooxanthellae) and their distribution and abundance at Lizard Island (Great
797 Barrier Reef). *Journal of Experimental Marine Biology and Ecology*, *364*(1), 48–53.
798 https://doi.org/10.1016/j.jembe.2008.06.034
799 Liu, H., Stephens, T. G., González-Pech, R. A., Beltran, V. H., Lapeyre, B., Bongaerts, P., Cooke,
800 I., Aranda, M., Bourne, D. G., Forêt, S., Miller, D. J., van Oppen, M. J. H., Voolstra, C. R.,
801 Ragan, M. A., & Chan, C. X. (2018). *Symbiodinium* genomes reveal adaptive evolution of
802 functions related to coral-dinoflagellate symbiosis. *Communications Biology*, *1*(1), 95.
803 https://doi.org/10.1038/s42003-018-0098-3
804 Matthews, J. L., Raina, J., Kahlke, T., Seymour, J. R., Oppen, M. J. H., & Suggett, D. J. (2020).
805 Symbiodiniaceae-bacteria interactions: rethinking metabolite exchange in reef-building corals as
806 multi-partner metabolic networks. *Environmental Microbiology*, *22*(5), 1675–1687.
807 https://doi.org/10.1111/1462-2920.14918

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.11.487905; this version posted April 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

808 Messyasz, A., Rosales, S. M., Mueller, R. S., Sawyer, T., Correa, A. M. S., Thurber, A. R., & Vega
809 Thurber, R. (2020). Coral bleaching phenotypes associated with differential abundances of
810 Nucleocytoplasmic Large DNA Viruses. *Frontiers in Marine Science*, *7*, 555474.
811 https://doi.org/10.3389/fmars.2020.555474
812 Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A.,
813 & Lanfear, R. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference
814 in the genomic era. *Molecular Biology and Evolution*, *37*(5), 1530–1534.
815 https://doi.org/10.1093/molbev/msaa015
816 Mita, P., & Boeke, J. D. (2016). How retrotransposons shape genome regulation. *Current Opinion*
817 *in Genetics & Development*, *37*, 90–100. https://doi.org/10.1016/j.gde.2016.01.001
818 Moniruzzaman, M., Weinheimer, A. R., Martinez-Gutierrez, C. A., & Aylward, F. O. (2020).
819 Widespread endogenization of giant viruses shapes genomes of green algae. *Nature*, *588*(7836),
820 141–145. https://doi.org/10.1038/s41586-020-2924-2
821 Montalvo-Proaño, J., Buerger, P., Weynberg, K. D., & van Oppen, M. J. H. (2017). A PCR-Based
822 Assay targeting the major capsid protein gene of a Dinorna-Like ssRNA virus that infects coral
823 photosymbionts. *Frontiers in Microbiology*, *8*, 1665. https://doi.org/10.3389/fmicb.2017.01665
824 Muller, E. M., Bartels, E., & Baums, I. B. (2018). Bleaching causes loss of disease resistance
825 within the threatened coral species Acropora cervicornis. *ELife*, *7*, e35066.
826 https://doi.org/10.7554/eLife.35066
827 Nagasaki, K., Shirai, Y., Takao, Y., Mizumoto, H., Nishida, K., & Tomaru, Y. (2005). Comparison
828 of genome sequences of single-stranded RNA viruses infecting the bivalve-killing dinoflagellate
829 *Heterocapsa circularisquama*. *Applied and Environmental Microbiology*, *71*(12), 8888–8894.
830 https://doi.org/10.1128/AEM.71.12.8888-8894.2005

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.11.487905; this version posted April 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

831 Nand, A., Zhan, Y., Salazar, O. R., Aranda, M., Voolstra, C. R., & Dekker, J. (2021). Genetic and
832 spatial organization of the unusual chromosomes of the dinoflagellate *Symbiodinium*
833 *microadriaticum*. *Nature Genetics*, *53*(5), 618–629. https://doi.org/10.1038/s41588-021-00841-y
834 Oliveira, N. M., Satija, H., Kouwenhoven, I. A., & Eiden, M. V. (2007). Changes in viral protein
835 function that accompany retroviral endogenization. *Proceedings of the National Academy of*
836 *Sciences*, *104*(44), 17506–17511. https://doi.org/10.1073/pnas.0704313104
837 Palatini, U., Miesen, P., Carballar-Lejarazu, R., Ometto, L., Rizzo, E., Tu, Z., van Rij, R. P., &
838 Bonizzoni, M. (2017). Comparative genomics shows that viral integrations are abundant and
839 express piRNAs in the arboviral vectors *Aedes aegypti* and *Aedes albopictus*. *BMC Genomics*,
840 *18*(1), 512. https://doi.org/10.1186/s12864-017-3903-3
841 Parker, B. J., & Brisson, J. A. (2019). A laterally transferred viral gene modifies aphid wing
842 plasticity. *Current Biology*, *29*(12), 2098-2103.e5. https://doi.org/10.1016/j.cub.2019.05.041
843 Patel, M. R., Emerman, M., & Malik, H. S. (2011). Paleovirology—ghosts and gifts of viruses past.
844 *Current Opinion in Virology*, *1*(4), 304–309. https://doi.org/10.1016/j.coviro.2011.06.007
845 Pesant, Stéphane, Lombard, Fabien, Bourdin, Guillaume, Poulain, Julie, Petit, Emmanuelle, Boss,
846 Emmanuel, Cassar, Nicolas, Cohen, Natalie R., Dimier, Céline, Douville, Eric, Flores, J. Michel,
847 Gorsky, Gabriel, Hume, Benjamin C.C., John, Seth G., Kelly, Rachel L., Lin, Yajuan, Marie,
848 Dominique, Pedrotti, Maria-Luiza, Pujo Pay, Mireille, … Allemand, Denis. (2020). *Tara Pacific*
849 *samples provenance and environmental context - version 1*. Zenodo.
850 https://doi.org/10.5281/ZENODO.4068293
851 Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic
852 features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.11.487905; this version posted April 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

853 Robbins, S. J., ReFuGe2020 Consortium, Singleton, C. M., Chan, C. X., Messer, L. F., Geers, A.
854 U., Ying, H., Baker, A., Bell, S. C., Morrow, K. M., Ragan, M. A., Miller, D. J., Forêt, S.,
855 Voolstra, C. R., Tyson, G. W., & Bourne, D. G. (2019). A genomic view of the reef-building
856 coral *Porites lutea* and its microbial symbionts. *Nature Microbiology*, *4*(12), 2090–2100.
857 https://doi.org/10.1038/s41564-019-0532-4
858 Roossinck, M. J. (2011). The good viruses: viral mutualistic symbioses. *Nature Reviews*
859 *Microbiology*, *9*(2), 99–108. https://doi.org/10.1038/nrmicro2491
860 Scheufen, T., Iglesias-Prieto, R., & Enríquez, S. (2017). Changes in the number of symbionts and
861 *Symbiodinium* cell pigmentation modulate differentially coral light absorption and
862 photosynthetic performance. *Frontiers in Marine Science*, *4*, 309.
863 https://doi.org/10.3389/fmars.2017.00309
864 Shoguchi, E., Beedessee, G., Hisata, K., Tada, I., Narisoko, H., Satoh, N., Kawachi, M., &
865 Shinzato, C. (2021). A new dinoflagellate genome illuminates a conserved gene cluster involved
866 in sunscreen biosynthesis. *Genome Biology and Evolution*, *13*(2), evaa235.
867 https://doi.org/10.1093/gbe/evaa235
868 Shoguchi, E., Beedessee, G., Tada, I., Hisata, K., Kawashima, T., Takeuchi, T., Arakaki, N., Fujie,
869 M., Koyanagi, R., Roy, M. C., Kawachi, M., Hidaka, M., Satoh, N., & Shinzato, C. (2018). Two
870 divergent *Symbiodinium* genomes reveal conservation of a gene cluster for sunscreen
871 biosynthesis and recently lost genes. *BMC Genomics*, *19*(1), 458.
872 https://doi.org/10.1186/s12864-018-4857-9
873 Shoguchi, E., Shinzato, C., Kawashima, T., Gyoja, F., Mungpakdee, S., Koyanagi, R., Takeuchi,
874 T., Hisata, K., Tanaka, M., Fujiwara, M., Hamada, M., Seidi, A., Fujie, M., Usami, T., Goto, H.,
875 Yamasaki, S., Arakaki, N., Suzuki, Y., Sugano, S., … Satoh, N. (2013). Draft assembly of the

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.11.487905; this version posted April 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

876 *Symbiodinium minutum* nuclear genome reveals dinoflagellate gene structure. *Current Biology*,
877 *23*(15), 1399–1408. https://doi.org/10.1016/j.cub.2013.05.062
878 Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H.,
879 Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of
880 high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems*
881 *Biology*, *7*(1), 539. https://doi.org/10.1038/msb.2011.75
882 Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015).
883 BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.
884 *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351
885 Skirmuntt, E. C., Escalera-Zamudio, M., Teeling, E. C., Smith, A., & Katzourakis, A. (2020). The
886 potential role of Endogenous Viral Elements in the evolution of bats as reservoirs for zoonotic
887 viruses. *Annual Review of Virology*, *7*(1), 103–119. https://doi.org/10.1146/annurev-virology-
888 092818-015613
889 Sofuku, K., & Honda, T. (2018). *Influence of Endogenous Viral Sequences on Gene Expression*.
890 IntechOpen. https://doi.org/10.5772/intechopen.71864
891 Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., & Morgenstern, B. (2006). AUGUSTUS:
892 ab initio prediction of alternative transcripts. *Nucleic Acids Research*, *34*(Web Server), W435–
893 W439. https://doi.org/10.1093/nar/gkl200
894 Stephens, T. G., González-Pech, R. A., Cheng, Y., Mohamed, A. R., Burt, D. W., Bhattacharya, D.,
895 Ragan, M. A., & Chan, C. X. (2020). Genomes of the dinoflagellate *Polarella glacialis* encode
896 tandemly repeated single-exon genes with adaptive functions. *BMC Biology*, *18*(1), 56.
897 https://doi.org/10.1186/s12915-020-00782-8

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.11.487905; this version posted April 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

898 Stoye, J. P. (2012). Studies of endogenous retroviruses reveal a continuing evolutionary saga.
899 *Nature Reviews Microbiology*, *10*(6), 395–406. https://doi.org/10.1038/nrmicro2783
900 Suzuki, Y., Baidaliuk, A., Miesen, P., Frangeul, L., Crist, A. B., Merkling, S. H., Fontaine, A.,
901 Lequime, S., Moltini-Conclois, I., Blanc, H., van Rij, R. P., Lambrechts, L., & Saleh, M.-C.
902 (2020). Non-retroviral Endogenous Viral Element limits cognate virus replication in *Aedes*
903 *aegypti* ovaries. *Current Biology*, *30*(18), 3495-3506.e6.
904 https://doi.org/10.1016/j.cub.2020.06.057
905 Takahashi, H., Fukuhara, T., Kitazawa, H., & Kormelink, R. (2019). Virus latency and the impact
906 on plants. *Frontiers in Microbiology*, *10*, 2764. https://doi.org/10.3389/fmicb.2019.02764
907 Tara Oceans Consortium Coordinators, Pesant, S., Not, F., Picheral, M., Kandels-Lewis, S., Le
908 Bescot, N., Gorsky, G., Iudicone, D., Karsenti, E., Speich, S., Troublé, R., Dimier, C., &
909 Searson, S. (2015). Open science resources for the discovery and analysis of Tara Oceans data.
910 *Scientific Data*, *2*(1), 150023. https://doi.org/10.1038/sdata.2015.23
911 ter Horst, A. M., Nigg, J. C., Dekker, F. M., & Falk, B. W. (n.d.). Endogenous Viral Elements are
912 widespread in arthropod genomes and commonly give rise to PIWI-interacting RNAs. *Journal of*
913 *Virology*, *93*(6), e02124-18. https://doi.org/10.1128/JVI.02124-18
914 Vega Thurber, R., Payet, J. P., Thurber, A. R., & Correa, A. M. S. (2017). Virus–host interactions
915 and their roles in coral reef health and disease. *Nature Reviews Microbiology*, *15*(4), 205–216.
916 https://doi.org/10.1038/nrmicro.2016.176
917 Tromas, N., Zwart, M. P., Forment, J., & Elena, S. F. (2014). Shrinkage of genome size in a plant
918 RNA virus upon transfer of an essential viral gene into the host genome. *Genome Biology and*
919 *Evolution*, *6*(3), 538–550. https://doi.org/10.1093/gbe/evu036

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.11.487905; this version posted April 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

van Oppen, M. J. H., Leong, J.-A., & Gates, R. D. (2009). Coral-virus interactions: A double-edged sword? *Symbiosis*, *47*(1), 1–8. https://doi.org/10.1007/BF03179964

Wang, L., Wu, S., Liu, T., Sun, J., Chi, S., Liu, C., Li, X., Yin, J., Wang, X., & Yu, J. (2014). Endogenous viral elements in algal genomes. *Acta Oceanologica Sinica*, *33*(2), 102–107. https://doi.org/10.1007/s13131-014-0446-1

Weynberg, K. D., Wood-Charlson, E. M., Suttle, C. A., & van Oppen, M. J. H. (2014). Generating viral metagenomes from the coral holobiont. *Frontiers in Microbiology*, *5*. https://doi.org/10.3389/fmicb.2014.00206

Whitfield, Z. J., Dolan, P. T., Kunitomi, M., Tassetto, M., Seetin, M. G., Oh, S., Heiner, C., Paxinos, E., & Andino, R. (2017). The diversity, structure, and function of heritable adaptive immunity sequences in the *Aedes aegypti* genome. *Current Biology*, *27*(22), 3511-3519.e7. https://doi.org/10.1016/j.cub.2017.09.067

Wilson, W., Francis, I., Ryan, K., & Davy, S. (2001). Temperature induction of viruses in symbiotic dinoflagellates. *Aquatic Microbial Ecology*, *25*, 99–102. https://doi.org/10.3354/ame025099

Wood-Charlson, E. M., Weynberg, K. D., Suttle, C. A., Roux, S., & van Oppen, M. J. H. (2015). Metagenomic characterization of viral communities in corals: mining biological signal from methodological noise: Methodological biases in coral viromics. *Environmental Microbiology*, *17*(10), 3440–3449. https://doi.org/10.1111/1462-2920.12803

Zhang, H., Zhuang, Y., Gill, J., & Lin, S. (2013). Proof that dinoflagellate Spliced Leader (DinoSL) is a useful hook for fishing dinoflagellate transcripts from mixed microbial samples: *Symbiodinium kawagutii* as a case study. *Protist*, *164*(4), 510–527. https://doi.org/10.1016/j.protis.2013.04.002

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.11.487905; this version posted April 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

943 Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic*

944 *Acids Research*, *31*(13), 3406–3415. https://doi.org/10.1093/nar/gkg595