Reef Invertebrate Viromics: Diversity, Host-Specificity & Functional Capacity

Patrick W. Laffy*[1], Elisha M. Wood-Charlson[§1], Dmitrij Turaev[2], Sabrina Jutz[2], Cecilia Pascelli[1,3,4], Emmanuelle S. Botté[1], Sara C. Bell[1], Tyler Peirce[1], Karen D. Weynberg[1], Madeleine J.H van Oppen[1, 5], Thomas Rattei[2] and Nicole S. Webster[1,6]

[1] Australian Institute of Marine Science, PMB 3, Townsville 4810, Queensland, Australia

[2] Department of Microbiology and Ecosystem Science, Division of Computational Systems Biology, University of Vienna, Vienna, Austria

[3]College of Science and Engineering, James Cook University, Townsville, Queensland, Australia.

[4]AIMS@JCU, Australian Institute of Marine Science and James Cook University, Townsville, Queensland, Australia.

[5]School of Biosciences, University of Melbourne, Parkville, Melbourne 3010, Victoria, Australia

[6]Austalian Centre for Ecogenomics, University of Queensland, Brisbane 4072, Queensland, Australia

[§]Present address: Center for Microbial Oceanography: Research and Education, University of Hawai'i at Mānoa, Hawai'i, USA

Corresponding Author*

Patrick W. Laffy
Australian Institute of Marine Science
PMB No 3, Townsville MC,
Townsville City, Queensland 4810 Australia
Email: p.laffy@aims.gov.au
Ph +61 7 4753 4365
Fax +61 7 4753 5852

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of record. Please cite this article as doi:10.1111/1462-2920.14110.

Originality-Significance Statement

The key elements of novelty of this work which places this work within the top 10% of current research are highlighted hereafter:

- We present the first taxonomic and functional description of viruses associated with marine sponges and **present a meta-analysis of viruses associated with sponges, corals and seawater** to describe variability in community dynamics across different niches of the reef environment.
- We provide **an assessment of viral taxonomic composition and OTU prevalence in host-associated systems** using a novel viral orthologous group-based OTU approach.
- We show that **viruses in reef invertebrates are host-specific and have functional repertoires that are distinct from viruses in the surrounding seawater**. The presence of auxiliary genes involved in herbicide resistance and specific viral pathogenesis pathways demonstrate the unique viral adaptations to their specific host microenvironments.
- We show that **the abundance of ssDNA viruses in reef invertebrates is comparable to dsDNA phages**, and identify previously unreported members of the ssDNA viruses within marine invertebrates such as the *Bidnaviridae* family.

Summary

Recent metagenomic analyses have revealed a high diversity of viruses in the pelagic ocean and uncovered clear habitat-specific viral distribution patterns. Conversely, similar insights into the composition, host-specificity and function of viruses associated with marine organisms have been limited by challenges associated with sampling and computational analysis. Here we performed targeted viromic analysis of six coral reef invertebrate species and their surrounding seawater to deliver taxonomic and functional profiles of viruses associated with reef organisms. Sponges and corals host species-specific viral assemblages with low sequence identity to known viral genomes. While core viral genes involved in capsid formation, tail structure and infection mechanisms were observed across all reef samples, auxiliary genes including those involved in herbicide resistance and viral pathogenesis pathways such as host immune suppression were differentially enriched in reef hosts. Utilising a novel OTU based assessment, we also show a prevalence of dsDNA viruses belonging to the *Mimiviridae, Caudovirales* and *Phycodnaviridae* in reef environments and further highlight the abundance of ssDNA viruses belonging to the *Circoviridae*, *Parvoviridae, Bidnaviridae* and *Microviridae* in reef invertebrates. These insights into coral reef viruses provide an important framework for future research into how viruses contribute to the health and evolution of reef organisms.

Introduction

Marine invertebrates are at the core of healthy coral reef ecosystems. Scleractinian corals are the fundamental ecosystem engineers, creating large intricate reefs that support diverse and abundant marine life, and filter feeders such as sponges process large volumes of seawater, recycle essential nutrients and play roles in the erosion and stabilization of coral reef structure (Ruetzler, 2004; Bell, 2008). Each individual reef invertebrate hosts a diverse micro-ecosystem comprised of bacteria, archaea, fungi, protists and viruses (Kushmaro et al., 2001; Bourne et al., 2016; Webster and Thomas, 2016; Welsh et al., 2016). These microbial communities can be host species-specific and extremely diverse, with thousands of unique microbial operational taxonomic units (OTUs) known to occupy some reef species (Thomas et al., 2016). While the complex microbial assemblages of reef invertebrates can be maintained across spatial, temporal and environmental scales (Taylor et al., 2013), comparatively little is known about the composition or variability of the associated viral communities. Similarly, while reef invertebrates are known to rely on their microbial communities for nitrogen, carbon and sulfur cycling, secondary metabolite production, and uptake and conversion of dissolved organic matter (Lesser et al., 2007; Taylor et al., 2007; Kimes et al., 2010; Fan et al., 2012; Ceh et al., 2013), insights into how viruses contribute to holobiont function have been elusive. The Seed-bank hypothesis, which identified a high-local and low-global diversity of viruses (Breitbart and Rohwer, 2005) was initially validated by analysis of virome communities from four oceanic regions (Angly et al., 2006) and was subsequently confirmed by large scale metagenomic analyses of epipelagic and mesopelagic seawater (Brum et al., 2015; Roux et al., 2016). These analyses identified that local viral community composition can shift in response to changing environment conditions.

Virus like particles (VLPs) were first observed in sponges by transmission electron microscopy (TEM) in 1977 (Vacelet and Donadey, 1977) but it was almost 30 years later that microscopy first revealed virus-like particles in the corals *Pavona danai* (Wilson et al., 2005), *Acropora formosa* (Davy and Patten, 2007) and *Acropora muricata* (Patten et al., 2008). The subsequent development of novel viral isolation methods (Marhaver et al., 2008; Thurber et al., 2009; Weynberg et al., 2014) facilitated the sequencing of viromes from eight scleractinian species (Wegley et al., 2007; Marhaver et al., 2008; Vega Thurber et al., 2008; Littman et al., 2011; Correa et al., 2013; Soffer et al., 2014; Weynberg et al., 2014; Correa et al., 2016), revealing that corals associate with up to 60 different viral families (Wood-Charlson et al., 2015). A recent review suggested that 9-12 of these viral families make up the 'core' coral virome (Vega Thurber et al., 2017), including the bacteriophage families *Myoviridae, Podoviridae* and *Siphoviridae* and the eukaryote-infecting nucleo-cytoplasmic large DNA virus (NCLDV) families *Phycodnaviridae* and *Mimiviridae* (Wood-Charlson et al., 2015; Correa et al., 2016; Laffy et al., 2016). Viruses matching the *Poxviridae, Iridoviridae* and *Ascoviridae*

families have also been found in more than 90% of all coral viromes (Wood-Charlson et al., 2015; Vega Thurber et al., 2017). The frequency of viruses matching *Herpesviridae* is more variable across coral species (Wood-Charlson et al., 2015), but this family has been observed in healthy and bleached corals using both TEM and metagenomics (Correa et al., 2016). Furthermore, *Phycodnaviridae* have been reported from both transcriptomic (Levin et al., 2017) and viromic (Correa et al., 2013; Weynberg et al., 2017) datasets derived from coral-isolated *Symbiodinium* cultures. ssDNA viruses have also been proposed as members of the core coral virome (Vega Thurber et al., 2017), with the prevalence of *Circoviridae* and *Microviridae* increasing in stressed/ bleached corals (Littman et al., 2011; Soffer et al., 2014). The ssRNA Retrotranscribing family *Retroviridae* and the dsDNA retrotranscribing family *Caulimoviridae* have also been proposed as members of the core coral virome (Vega Thurber et al., 2017). In comparison, there has only been a single study characterising sponge-associated viruses which reported a predominance of *Caudovirales* as well as members of the *Mimiviridae*, *Phycodnaviridae* and ssDNA viruses (Laffy et al., 2016).

Given that viruses can be responsible for reprogramming host metabolism, mediating lateral gene transfer and structuring microbial communities, there is clearly a need to ascertain the functional roles these viruses are playing within reef holobionts. A landmark study of environmental viromes by Dinsdale and colleagues in 2008 used a SEED Subsystems approach to provide the first functional assessment of coral-associated viruses (Dinsdale et al., 2008). However, most subsequent studies focussed on community characterisation (Wegley et al., 2007; Marhaver et al., 2008; Vega Thurber et al., 2008; Littman et al., 2011; Correa et al., 2013; Soffer et al., 2014; Weynberg et al., 2014; Correa et al., 2016), with limited additional insights into the functional potential of coral reef viruses. A recently developed computational workflow specifically designed for holobiont-associated viromes (Laffy et al., 2016), incorporates an analyses of viral function based on Swiss-Prot Keyword associations derived from pairwise similarity matches. We utilised this workflow to compare the taxonomy and function of viruses inhabiting four Great Barrier Reef (GBR) sponge species (*Amphimedon queenslandica*, *Xestospongia testudinaria*, *Ianthella basta* and *Rhopaloeides odorabile*), two coral species collected from the GBR (*Pocillopora acuta*) and Papua New Guinea (PNG) (*Porites lutea*) and seawater collected from both locations. We also test whether the spatially structured viral diversity trends identified in a global analysis of the Earth's virome (Brum et al., 2015) extend to fine spatial scales such as individual reef species.

Results and Discussion

Sequence data evaluation

In total, over 52 million paired end 250bp Illumina reads which generated 96,757 contigs containing 300,783 predicted genes were analysed (Supp.Table S1). Single read analysis following the HoloVir protocol facilitated taxonomic identification of between 2.2% to 26.3% of reads from each dataset, and assembled gene analysis resulted in taxonomic identification of 26.2% to 58.1% of all assembled contigs (Supp.Table S1). Between 4.8-13.8% of the predicted genes were assigned function based on Swissprot keyword identification and 68.7-98.5% of the source contigs with functional assignments were also assigned viral taxonomies (Supp.Table S1).

Evaluation of single read data identified that between 0 and 0.2% of all reads showed significant matches to rRNA genes, and following the convention described by Roux and colleagues for evaluating cellular contamination in virome datasets (Roux et al., 2013), the majority of samples could therefore be categorised as having no cellular sequences, or very low to negligible levels of cellular contamination (Supp Table S1). Samples *X. testudinaria* 1 and 3, *I. basta* 3, *R. odorabile* 4, *P. acuta* 1 and 3 and all *P. lutea* samples contained slightly higher levels of rRNA, therefore these samples were categorised as having non-negligible proportions of cellular sequences (Roux et al., 2013) (Supp Table S1). The cellular marker gene analysis showed that all datasets which had no, low or negligible rRNA gene matches according to Roux and colleagues also had fewer than 0.1% cellular marker matches, with the exception of *X. testutinaria* 2, which had 0.3% cellular marker matches. Datasets identified as having non-negligble cellular contamination contained up to 0.3% cellular marker matches (Supp Table S1). However, assembly further reduced cellular contamination in these samples with no more than 6 contigs having matches to rRNA genes within SILVA and no more than 11 contigs across the entire dataset having matches to HoloVir cellular markers (<0.2% of all contigs) (Supp Table S1). This low frequency of rRNA matches demonstrates the utility of the assembly strategy for minimising the number of contigs derived from cellular contamination.

Community structure analysis

Analysis of the viromes derived from corals and sponges revealed highly conserved viral communities amongst the 3-5 biological replicates of each host species (with the exception of the coral, *P. acuta*), with viral communities being clearly distinct from those inhabiting the surrounding seawater (Figure 1). The composition of the viral community was significantly different amongst host species (PERMANOVA, Pseudo-F_7 = 7.3, p=0.001) and host environments (i.e. coral vs sponge vs seawater) (PERMANOVA, Pseudo-F_3 = 3.4, p=0.001). All species hosted viral communities that were significantly different from each other, with the exception of *A. queenslandica*, for which the viral community was only significantly different to that of *R. odorabile*, *P. lutea* and PNG seawater

samples (Table 1). Sponge viromes exhibited higher intra-species similarity (78-86%) than coral viromes (66-76%), which is remarkable considering sponges filter thousands of litres of seawater each day, efficiently extracting and digesting the virioplankton and bacterioplankton (Ludeman et al., 2016). Microbial community dynamics in reef corals and sponges has been studied extensively (Bourne et al 2016, Webster and Thomas, 2016), with a recent global analysis of the sponge microbiome revealing high variability in microbial richness across different host species but low intra-species microbial variability (Thomas et al., 2016). Similarly, the human gut microbiome and virome have been shown to exhibit minimal temporal variability within individuals but greater interpersonal variability (Costello et al., 2009; Reyes et al., 2010). The low intra-species variability in the viral communities inhabiting the four sponge species is consistent with the previously described low microbial variability in these species (Ruetzler, 2004; Fan et al., 2012; Luter et al., 2012; Gauthier et al., 2016). In contrast, the microbiome of some coral species (including *P. lutea*) has been reported to vary both spatially and temporally (Sunagawa et al., 2009; Sunagawa et al., 2010; Littman et al., 2011; Morrow et al., 2012; Li et al., 2014; Morrow et al., 2014).

A large scale investigation into global ocean viromes identified that nutrient concentrations influence viral community structure at specific geographic locations, and that environmental conditions directly influence global viral distribution (Brum et al., 2015). Similarly, the significantly different viral community structures we observed in corals and sponges are likely influenced by the varied nutrient and metabolite concentrations of their specific host-associated habitats (Fiore et al., 2010; Cachet et al., 2015; Sogin et al., 2017). The global ocean virome analysis also reported that less than one percent of all ocean viruses were affiliated to cultured representatives (Brum et al., 2015), likely explaining the high frequency of unidentified viral contigs within our coral and sponge-associated viral datasets (Supp Table S2).

Comparative analysis of viral taxonomy

Assessment of viral community composition revealed that the *Caudovirales* (dsDNA) and *Microviridae* (ssDNA) were the most diverse taxa across all reef invertebrates, with 100s of unique viral OTUs per host species (Figure 2). The Order *Caudovirales* had the highest number of individual OTUs in all samples, although variation in the three *Caudovirales* families was evident across invertebrate species, with *I. basta*, *A. queenslandica* and *R. odorabile* being dominated by *Myoviridae*, *Siphoviridae* and *Podoviridae* respectively. The abundance of individual viral taxa likely correlates with the abundance of specific symbiotic microorganisms in each species. For instance, the microbiome of *A. queenslandica* is known to be dominated by a sulfur-oxidising *Gammaproteobacteria* and a *Betaproteobacteria*, both of which are known to be targeted by siphoviruses (King et al., 2011). Importantly however, while there is a high level of similarity in the

microbiomes of *X. testudinaria* and *R. odorabile* (Thomas et al., 2016), their viral communities were significantly different, highlighting that viral composition cannot be inferred across holobiont host species based solely on similarity in bacterial community composition (Sullivan et al., 2003; Dekel-Bird et al., 2015). The ssDNA *Microviridae*, which are known to infect *Enterobacteriaceae* and *Tenericutes*, were highly diverse and prominent in all reef invertebrate taxa, although their representation was greatly reduced in PNG seawater samples (Figure 2). Importantly, while only five known Microvirus genera could be discerned (Tables S3-S6), the number of *Microviridae* OTUs was much higher, indicating considerable strain-level variation in the *Microviridae* or the presence of previously undocumented ssDNA microvirus species. Up to 40 *Circoviridae* OTUs were also evident across the GBR sponge, coral and seawater samples. The presence of ssDNA Microviruses in coral virome data sets has been reported elsewhere (reviewed in (Wood-Charlson et al., 2015)), although it was previously assumed their prominence was due to Phi29-based multiple displacement amplification methods preferentially enriching ssDNA molecules (Wood-Charlson et al., 2015). Members of the Microviridae have also been identified in the human gut, where they are responsible for increased variability in comparison to temperate phages (Minot et al., 2013). The abundance of this viral family shown here and also recently reported in global ocean surveys (Labonté and Suttle, 2013; Székely and Breitbart, 2016), highlights the ubiquity and importance of *Microviridae* in marine environments.

Representatives from the *Megavirales* families *Ascoviridae*, *Iridoviridae*, *Poxviridae*, and *Marseilleviridae* (Colson et al., 2013) were observed in individual replicates across multiple species (Figure 2), although the number of observed OTUs was generally low. Representatives of the *Megavirales* are known to target a wide range of eukaryotic hosts, indicating that these viruses are specifically targeting the coral / sponge cells or their co-occurring eukaryotic symbionts, which would extend their known host range, or that they are infecting the invertebrate-associated meiofauna. Interestingly, a low number of OTUs assigned to the *Mimiviridae*, which are known to infect amoeba (La Scola et al., 2003), was detected across all reef environments. Sponges have mobile amoebocyte cells which move throughout the sponge body and likely play host to these novel *Mimiviridae* (Claverie, 2009). Pairwise sequence comparison of all *Mimiviridae* genes from each dataset confirmed that few homologous genes occurred across replicate samples, with the exception of the *I. basta* biological replicates, where several *Mimiviridae* genes shared > 70% sequence homology (Supp. Figure S1). Additionally, almost no *Mimiviridae* sequence identity was observed between sample types, suggesting there are undescribed *Mimiviridae* in reef invertebrates which are host species-specific (Supp. Figure S1). Members of the *Herpesviridae* were largely absent from all samples analysed, but a single OTU was detected in low abundance in a single sample of *A.*

queenslandica (Figure 2). Herpes-like viruses have previously been reported to infect bleached, diseased and stressed corals (Vega Thurber et al., 2008; Soffer et al., 2014; Correa et al., 2016) but were not identified in any of the coral samples from the present study, possibly reflecting species-specific, health state or methodological differences (Weynberg et al., 2014; Wood-Charlson et al., 2015).

Algae-infecting *Phycodnaviridae* were also detected across all sample types, being particularly prevalent in the sponges' *A. queenslandica* and *I. basta* which are known to host low microbial diversity and no photosymbionts (Luter et al., 2010; Gauthier et al., 2016). The absence of a known *Phycodnaviridae* host within the sponge holobionts indicates either an expansion of the host range for this viral family or that these particular sponge species filter and concentrate phototrophic microorganisms from the seawater. In corals, the *Phycodnaviridae* likely target the associated dinoflagellate symbionts *Symbiodinium* (Correa et al., 2016; Levin et al., 2017; Weynberg et al., 2017). *Parvoviridae*, which were traditionally thought to infect vertebrates and arthropods but have recently been reported from a range of marine invertebrates including corals (François et al., 2016), were particularly prevalent in the sponges *I. basta, R. odorabile* and the coral *P. acuta* as well as the GBR seawater. Here we broaden the host range of *Parvoviridae* to include the *Porifera* holobiont, where they occurred in three of the four sponge species. *Iridoviridae*, which are known to infect invertebrates (Williams, 2008) were detected in low abundance in *A. queenslandica* and *P. lutea*. *Poxviridae,* which are known to infect insect, arthropod and vertebrate hosts (Hughes et al., 2010; Haller et al., 2014), were detected in some replicates of *A. queenslandica*, *I. basta, R. odorabile* and *P. lutea*. ssDNA viruses from the family *Bidnaviridae* were additionally detected in the viral communities of *R. odorabile*, *I. basta* and *P. lutea* using last common ancestor based viral RefSeq taxonomic assignment (Figures S2, S3, Table S2). Members of the *Bidnaviridae* are eukaryotic ssDNA viruses which were previously assigned to the *Parvoviridae* family, but which have incorporated Polintoviral, Baculoviral and Reoviral genes (Krupovic and Koonin, 2014).

As viral genome size can vary greatly between different viral families, it is important to consider the identification of individual genes from whole genome analysis of virome communities. Members of the *Circoviridae* have as few as two coding genes in their genome (Rosario et al., 2017), members of the *Microviridae* typically have between three to nine genes (Roux et al., 2012), while members of the giant viruses in the family *Mimiviridae* can have more than 900 genes (Claverie, 2009). While OTU frequency calculations were normalised to account for the average number of marker genes that have been observed for all viral families (Figure 2), the abundance of taxonomic assignments for each viral family based on viral Refseq BLAST+ analysis does not take into account the variation in

genome size and gene frequency when assigning taxonomy. This likely overestimates the contribution of certain viral taxa in overall community composition (Figures S2, S3).

Comparative analysis of viral function

A total of 551 different Uniprot/Swissprot (further referred to as Swissprot) keywords were assigned to the predicted genes derived from coral, sponge and seawater viromes. In summary, 4.8%-13.8% of sponge-associated viral genes, 2.4%-29.1% of coral-associated viral genes and 3.2%-10.3% of seawater associated viral genes were assigned Swissprot keywords. When further investigating the 50 most common Swissprot keywords identified across all viromes, viral-specific functions dominated, comprising genes involved in i) viral capsid structure and formation (capsid proteins, plasmid partition, viral capsid assembly, virion and viral genome packaging), ii) viral tail structure (viral tail proteins, viral tail assembly and viral tail fibre proteins), iii) viral infection mechanisms (viral penetration into host cytoplasm, viral tail ejection systems, host cell lysis, bacteriolytic enzymes and DNA end degradation evasion) and iv) viral latency (viral latency and latency-replication switch) (Figure 3). A small number of additional genes were assigned non-viral Swissprot keywords including cobalamin, collagen, hydroxylation and elliptocytosis (Figure 3). The 50 most abundant Swissprot keywords were generally represented across multiple sample types (Figure 3), with the exception of viral latency genes, which were exclusively detected in *P. acuta*, and elliptocytosis, which was only observed in *P. lutea*. Genes involved in viral replication, including DNA polymerase, Primosome, and DNA replication, dominated most reef samples. A variety of different viral structural genes including capsid, tail fiber and baseplate proteins were also observed across most samples. Genes involved in specific viral pathogenesis pathways such as lipopolysaccharide degradation, viral host membrane-permeablization and evasion of bacteria-mediated translation shutoff were more variable across sample types (Figure 3), highlighting how different holobiont communities likely utilize varying infection and pathogenesis processes.

COG functional category assignment varied between sample types, with 11-17% of sponge-associated viral genes, 14-19% of coral-associated viral genes and 11-29% of seawater-associated viral genes being assigned COG functional categories. Over 75% of genes associated with the top 50 Swissprot keywords also had COG categories assigned as either unknown function (S) or could not be assigned into any COG (Figure 3), reflecting how COG functional categories were developed using bacterial rather than viral genomes and highlighting the superior functional resolution obtained using the Swissprot keyword enrichment approach (Galperin et al., 2015). Of the ~25% of genes assigned COG functional categories, most were designated as replication and repair (L), nucleotide metabolism and transport (F) and cell wall/membrane/envelope biogenesis (M) (Figure 3).

In addition to the use of marker gene and rRNA gene analysis to ensure assembled contigs were not significantly influenced by co-purified cellular contigs (Table S7), a detailed analysis of contigs coding for three functional keywords (collagen, herbicide resistance and cobalamin) was performed. In all cases, contigs containing these individual keywords were found on contigs containing viral taxonomic matches (Figure S4).

To determine which viral functions were significantly different across host species and habitats, we employed the multivariate statistical package mvabund (Wang et al., 2012) using the entire SwissProt functional keyword classification data set. Many genes involved in capsid and tail assembly and injection into target hosts were not significantly different across species or habitats. However, genes with keyword assignment to hydrolase, DNA repair, damage and binding, as well as DNA recombinase were found to be significantly enriched in both *A. queenslandica* and in the GBR seawater (Figure 4) with genes being assigned these keywords known to accumulate in viral genomes to facilitate efficient replication (Gillespie et al., 2012). Herbicide resistance genes and viral penetration via permeablization of host membranes were significantly enriched in *X. testudinaria* (Figure 4), and the contigs containing these genes were assigned to a *Synechococcus* phage belonging to the family *Myoviridae* (Table S7). It is likely that the viruses containing these auxiliary genes are targeting the dominant photosymbionts in *X. testudinaria*, the cyanobacterium *Synechococcus spongiarum* (Thomas et al., 2016). Herbicides are highly effective at controlling cyanobacterial populations (Perron and Juneau, 2011; Wu et al., 2016) and the introduction of viral herbicide resistance genes into plant genomes (de Freitas et al., 2007) has already been demonstrated. It is therefore conceivable that Myoviral cyanophages are providing herbicide resistance genes to their sponge-associated cyanobacterial hosts.

The Swissprot keywords exopolysaccharide synthesis, glycoprotein, calcium and enriched capsule biogenesis/degradation were significantly enriched in *I. basta* (Figure 4), and several of these keywords include genes involved in cell wall-viral interactions, although their exact role in *I. basta* remains to be determined. Peptidoglycan anchors were also enriched in *I. basta* (Figure 4), with the majority of contigs containing these genes assigned to *Caudovirales* (Table S7). Exposure to bacterial polysaccharides including peptidoglycan anchors has previously been shown to increase viral infectivity (Kuss et al., 2011), and hence these peptidoglycan anchors in *I. basta* derived *Caudovirales* may also be increasing infection rates into host cells.

Genes assigned the Swissprot keyword Cobalamin were detected in all samples but were significantly enriched in *I. basta* (Figure 4), and were taxonomically assigned to dsDNA viral contigs (Table S7). Cobalamin biosynthesis is undertaken by many bacteria and archaea, with Vitamin B12

being an essential cofactor required by most organisms (Doxey et al., 2014). Cobalamin biosynthesis genes have previously been identified in Myoviruses that infect *Prochlorococcus*, where they are suggested to boost cobalamin production during infection cycles to improve the activity of ribonucleotide reductase genes (Sullivan et al., 2005). *I. basta* is known to host populations of *Cyanobacteria* and the viral contig(s) containing the cobalamin biosynthesis genes may also be influencing ribonucleotide reductase genes in their hosts.

A significant enrichment of methyltransferase and chromatin regulator keywords were identified in *R. odorabile* (Figure 4). Methyltransferases have been identified in *Chlorella* viruses of the *Phycodnaviridae* family (Zhang et al., 1998) as well as in several dsDNA bacteriophages (Krabbe and Carlson, 1991), and are essential in highly methylated viral genomes (Nelson et al., 1993). As *R. odorabile* does not contain algal photosymbionts (Bourne et al., 2013), it is unlikely these genes originate from an algal virus. Chromatin regulation genes are critical for interaction with chromatin structure and formation, enabling infecting viruses access to specific genomic regions of their hosts (Lieberman, 2006). Genes involved in viral endocytosis by host were also enriched in *R. odorabile*, and these contigs were assigned to Parvoviruses of the subfamily *Densovirinae* (Table S7). Members of the *Densovirinae* are known to infect insects and arthropods (King et al., 2011), so their presence in *R. odorabile* indicates either an expansion of their host range into the *Porifera* or the presence of arthropods within the *R. odorabile* holobiont.

A significant enrichment of viral latency genes was evident in *P. acuta*, indicating an active viral infection. DNA end degradation evasion which enables viruses to actively avoid detection and degradation once they infect their target hosts were also significantly enriched in *P. acuta* (Dillingham and Kowalczykowski, 2008) (Figure 4). Contigs assigned to viral latency were found exclusively in *Siphoviridae* lambdalike phages (Table S7). Lambda phages are one of the few bacteriophage who have been identified to enter both a lytic and lysogenic life cycle (Salmond and Fineran, 2015). Genes assigned the Swissprot keyword nucleosome core were observed almost exclusively in *P. lutea* (Figure 4). These genes matched Histone H3 and originated from contigs assigned to giant viruses from the *Marseilleviridae* family (Table S7). Histones have previously been identified in *Marseilleviridae* (Erives, 2015), and are thought to be involved in compressing viral genomes and suppressing host cell immune reactions (Avgousti et al., 2016).

The increase in photosystem I genes observed in *P. lutea* (Figure 4) may indicate that viruses are playing a key role in the observed increase in photosystem genes within microbial communities at oligotrophic reef sites (Kelly et al., 2014). The presence of photosystem I and II genes in marine virus

genomes has been previously reported (Lindell et al., 2005), and have been proposed to increase viral fitness by supplementing host photosynthetic energy production (Lindell et al., 2005).

When differences between host environments were investigated, genes involved in hydroxylation, collagen and viral genome excision were found to be significantly enriched in sponge viromes and genes involved in viral latency, RNA directed DNA polymerases, transposable elements, toxins and formylation were significantly enriched in coral viromes (Figure 5). Collagen producing genes have been identified in nucleocytoplasmic large DNA viruses and as a structural component of the external capsids of mimiviruses (Shah et al., 2014), although none of the contigs containing collagen genes from this study were assigned to *Mimiviridae* (Table S7). While collagen is one of the most ubiquitous proteins produced in multicellular animals, it is also one of the main structural components which give form and structure to sponges (Simpson, 1984). While the specific functional role of viral collagen genes in reef holobionts remains to be determined, their abundance in different sponge species (Figure 5) may indicate that some viruses are specifically targeting collagen producing pathways within their hosts. Hydroxylation genes were also enriched in both the sponge and coral viromes, with potential roles ranging from antibiotic biosynthesis (Neary et al., 2007), regulation of transcription (Tsukada et al., 2006) and DNA repair (Trewick et al., 2002). Most of the genes significantly enriched in corals originated from *Caudovirales* contigs in *P. lutea* and *Siphoviridae* lambda phages in *P. acuta* (Table S7). In seawater samples, an enrichment of genes involved in helicase activity, viral short tail ejection systems and degradation of host chromosomes by viruses was observed (Figure 5).

Conclusion

Viromic analysis of reef sponges, corals and their surrounding seawater has generated valuable new insights into the ecology of viruses inhabiting reef ecosystems. Viruses infecting reef holobionts exhibit significant host species-specificity and appear uniquely adapted to their individual host habitats. Bacteriophage dominate the communities of corals and sponges, hence low variability in the viral assemblages likely reflects the low intra-species variability of the respective host microbiomes (Thomas et al., 2016). However, the finding that species with more similar microbial associations do not necessarily have more similar viral assemblages, suggests that the holobiont environment still allows for niche selection of viral communities. We also report the presence of ssDNA viruses of the *Bidnaviridae* family in sponges and corals, and further highlight the abundance of ssDNA viruses belonging to the *Circoviridae*, *Parvoviridae* and *Microviridae* in coral reef invertebrates, with the latter rivalling the number of OTUs found in the *Caudovirales* families. Many core viral functional genes were also conserved across all host environments, with genes involved in viral capsid formation, tail structure and infection mechanisms being observed in all reef samples.

Other functional genes were specifically enriched in individual holobiont species. For instance, lambda phage viral latency genes were exclusively detected in *P.lutea*, herbicide resistance genes were only enriched in *X. testudinaria*, cobalamin biosynthesis genes were only enriched in *R. odorabile* and DNA end degradation evasion genes were only enriched in *P. acuta*. These variable components likely enable viral assemblages to cope with different host defence strategies and may provide unique advantages to the holobiont host to increase community fitness. These insights into coral reef viral ecology provide the framework for future hypothesis driven research into how reef viruses contribute to holobiont health and evolution.

Experimental procedures

Sample collection

All coral and sponge samples were collected on SCUBA, at locations described in Table S1, and tissue was immediately processed for viral isolation or snap frozen for downstream processing (see below). Replicate tissue samples were collected within the same dive, limiting the distance between samples to a radius of 200 m. Papua New Guinea (PNG) (n= 4 x 12 L) seawater samples were collected in sterile containers, immediately pre-filtered using a 0.22 µm Sterivex polyethersulfone filter and the filtrate was concentrated via chemical flocculation as described in (John et al., 2011) and stored at 4°C on Sterivex filters until further processing. GBR (n= 3 x 30L) seawater samples were stored at 4°C for no more than 4 days before being concentrated to approximately 20 mL using Tangential Flow Filtration (30 kDa, Pall Corporation) as described in (Sun et al., 2014). PNG seawater samples were resuspended in ascorbate-EDTA buffer (John et al., 2011). Seawater samples were then concentrated using Amicon centrifugal spin columns (30kDa, Millipore) and washed with 0.02 µm filtered SM buffer before cesium chloride (CsCl) gradient separation as described in (Weynberg et al., 2014).

Sample homogenisation, cellular disruption and cesium chloride fractionation

For all coral samples, tissue was air blasted from the skeleton into 15 mL 0.02 µm filtered (Anotop, Whatman) SM buffer (100 mM NaCl, 8 mM $MgSO_4$, 50 mM Tris pH 7.5) in sterile zip-lock bags and processed as previously described (Weynberg et al., 2014). Tissue from *R.odorabile* and *I.basta* were snap frozen upon collection, while tissue from *X. testudinaria* and *A. queenslandica* was processed fresh. For subsequent processing all sponge species were cut into small pieces and homogenised in SM buffer for 10 min. Sponge homogenates were passed through a 100 µm sieve (Corning Life Sciences) and centrifuged at 500 g for 15 min. All sample supernatants were loaded onto CsCl density gradients and centrifuged as described previously (Weynberg et al., 2014). Gradient sities were determined gravimetrically and DNA concentrations were measured using a Quant-It Picogreen

dsDNA high sensitivity assay kit (Invitrogen, Live Technologies). Fractions containing nucleic acid concentrations greater than 100 ng/ml were pooled together prior to buffer exchange (to remove CsCl salts) using Amicon centrifugal spin columns (30kDa, Millipore) and 0.02 µm filtered SM buffer. The viscosity of the sponge samples necessitated 0.2 µm filtering prior to buffer exchange. After buffer exchange, all samples were filtered using 0.2 µm pore size Durapore® (low protein binding) syringe filters to remove any remaining cellular contamination.

Nucleic acid extraction, amplification and sequencing

All samples were treated with DNase (Ambion) and RNase (Ambion) for 30 mins at 37°C. Nucleic acids were extracted and RNase treated using a MasterPure™ DNA purification kit (Epicentre, Madison WI) (*P.lutea, P. acuta*, PNG seawater), UltraClean Microbial DNA isolation kit (MoBio, Carlsbad CA) (*I. basta*, *R. odorabile* and GBR seawater) or FastDNA™ SPIN Kit for Soil (MP Biomedicals, Santa Ana CA) (*A. queenslandica* and *X. testudinaria*) according to the Manufacturer's instructions (see Table S1 for full details). All seawater, coral and sponge samples were amplified using a modified Random Priming-mediated Sequence-Independent Single-Primer Amplification (SISPA) approach as previously described (Weynberg et al., 2014), to enrich nucleic acids prior to sequencing and convert all ssDNA sequences to dsDNA templates. Final amplified PCR products were cleaned using a MinElute® PCR purification kit and quantified using a Quant-IT PicoGreen® kit. PCR fragments were visualised on a 0.8% agarose gel. All purified viral metagenomes were sequenced using Nextera XT MiSeq 300 bp paired-end sequencing (Illumina) at the Ramaciotti Centre, University of New South Wales, Australia. All samples were submitted to the Genbank Sequence Read Archive as summarised in Table S1.

Sequence analysis

A full summary of the computational workflow can be seen in Figure 6. Briefly, sequence data sets were analysed using HoloVir, a computational workflow designed for assigning taxonomy and function to host-associated viromes (Laffy et al., 2016). HoloVir employs a two tiered approach, utilising single read data and assembled predicted gene data in viral RefSeq (release 74) pairwise sequence comparisons to perform taxonomic assignment, complemented by marker gene validation of taxonomy and assessment of cellular contamination. Functional categorization was performed on predicted genes by pairwise sequence comparisons to Uniprot/Swissprot protein database (release-2016 01) (The UniProt Consortium, 2015), and assigned to specific Clusters of Orthologous Groups based on comparisons to the EggNOG 4.5 database (Huerta-Cepas et al., 2015).

Single read analysis: Quality control evaluation, trimming and merging

Quality Control (QC) on raw sequence reads was performed using FastQC (version 0.11.5) (Andrews, 2010), overlapping reads from paired end data were identified and merged with PEAR (Zhang et al., 2014). Unmerged reads were fused together, separated by 10 padding n residues, to maintain the relationship between non-overlapping paired end reads without adversely influencing subsequent BLAST matches. Merged reads were dereplicated using cd-hit-est (Li and Godzik, 2006), with a global identity threshold of 99%. Merged and non-merged reads were combined prior to BLAST+ analysis for taxonomic assignment.

Single read analysis: Taxonomic assignment and validation

Comparison to the viral RefSeq Database (release 74)(Brister et al., 2015) via BLAST+ (Camacho et al., 2009) sequence similarity searches was performed using default parameters. Taxonomic assignments were performed using MEGAN5 (Huson et al., 2011) LCA default parameters, and a minimum support parameter of five reads. A cellular and viral marker database was generated as previously described (Laffy et al., 2016) and used in sequence similarity comparisons to confirm viral RefSeq taxonomic assignments and to identify potential cellular contaminants. Cellular contamination was also evaluated (Roux et al., 2013) by BLAST+ comparisons to the SILVA rRNA database (Release 128) (Quast et al., 2013).

OTU based estimates of viral community composition were calculated using VOGDB models (www.vogdb.org version 81) for four types of core viral proteins; helicase, terminase, polymerase and major capsid proteins. A total of 189 individual VOGS were combined across all four viral protein types, ensuring that at least one representative marker was included for all dsDNA, ssDNA and retro-transcribing viral families, including 32 major capsid protein VOGs, 35 Helicase VOGs, 104 polymerase VOGs and 18 terminase VOGs. Briefly, SymBets were used to construct viral intergenomic symmetrical best matches (Kristensen et al., 2010) before NCBI COGsoft was employed to iterate the edges of SymBet graphs (Kristensen et al., 2010). These were further clustered based on a HMM-HMM similarity graph to form the final viral orthologous group HMM models. MetaGeneAnnotator (Noguchi et al., 2008) was used to identify gene coding sequences from sequence reads and HMM-graspx (Zhong et al., 2016) was used to pre-assemble individual sequence reads to this cohort of 189 VOGDB models. Final assembly was performed by a combination of quality trimming using prinseq-lite (Schmieder and Edwards, 2011) using the following non-default parameters: -min_len 50 -min_qual_mean 20 -trim_qual_right 20 -trim_qual_window 10 -trim_qual_type min, and trinity was employed for assembly (using default parameters) (Haas et al., 2013) for each HMMer model. Reads were normalised to the same number of reads per sample,

using randomised subsampling, and total abundance of each viral family was calculated. OTU numbers were normalized by the number of marker families typically found in a taxon and used to calculate their abundance.

Gene-centric Analysis: Assembly, gene prediction and taxonomic assignment

De novo assembly of viral metagenomes was performed using CLC Genomics Workbench 8.5.1 (https://www.qiagenbioinformatics.com/) with subsequent filtering steps for a minimum of 3 x coverage and a minimum contig length of 1000 bp. Gene prediction was performed on assembled contigs using MetaGeneAnnotator (Noguchi et al., 2008). Predicted genes were screened and taxonomic assignments performed using the same approaches described for single read analysis, incorporating viral Refseq, HoloVir marker gene and rRNA comparisons. Bray-Curtis distance matrices were generated and Non-metric MultiDimensional Scaling (NMDS) analysis was used to visually compare virome communities. PERMANOVA (Primer6/PERMANOVA+ v1.0.2 ; Plymouth, UK) was performed on standardized square root transformed data using 999 permutations to test differences in community structure between species and habitat types (corals, sponges, seawater).

Gene-centric Analysis: Functional assignment

Predicted gene functions were determined by performing BLAST+ (Camacho et al., 2009) sequence similarity searches with an *e*–value cutoff of 10^{-10} against the complete UniprotKB/Swissprot functionally annotated database (The UniProt Consortium, 2015). SwissProt keywords were identified for each best hit and collated for each viral metagenome. The 50 most abundant keywords across all samples were identified and the predicted genes that gave rise to these keywords were identified. Predicted genes were similarly searched against the EggNOG 4.5 database (Huerta-Cepas et al., 2015) and where a significant match was observed, that is an *e*-value less than 10^{-10}, COG functional categories were assigned (Galperin et al., 2015). Genes which were responsible for the 50 most abundant keywords were cross referenced against the COG functional category classification (Figure 3). The R package mvabund (Wang et al., 2012) was used to perform univariate tests on permutational multivariate analysis of variance, identifying drivers of functional differences between species as well as between host environments.

In order to validate the viral origin of specific functional genes (herbicide resistance, cobalamin, collagen) source contigs were identified from each dataset and all predicted genes from each contig were submitted to nr BLASTP analysis. MEGAN5 LCA taxonomic assignment was performed using nr BLASTP results and overall taxonomy for each contig was summarised. Six taxonomic assignment categories were established; i) virus only, ii) virus and unresolved taxonomy, iii) virus and cellular, iv)

unresolved only, v) cellular and unresolved and vi) cellular only. The resulting nr taxonomic assignments are summarised in Figure S4.

Supplementary information is available at Environmental Microbiology website

Acknowledgements

The Authors declare no conflict of interest. NSW and MvO were funded through Australian Research Council Future Fellowships (FT120100480 and FT100100088 respectively), KDW was funded through an Australian Research Council Super Science Fellowship (FS110200034), CP was funded through a PhD Scholarship from The CAPES foundation and Science without Borders. Additional funding was provided from AIMS. The authors wish to acknowledge Hans-Jörg Hellinger for his contribution to development discussions on the VogDB analysis of virome datasets, and Hannah Epstein and Sam Noonan for providing photographs for the manuscript.

Author contributions

NSW, MvO, TR, KDW, EMWC and PWL conceived and designed the study. KDW, EMWC, ESB, SCB and CP undertook the laboratory work. TR and SJ developed the OTU based analysis. All remaining data analysis was performed by PWL, NSW, DR and TP. PWL and NSW wrote the manuscript.

References

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. URL http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Angly, F.E., Felts, B., Breitbart, M., Salamon, P., Edwards, R.A., Carlson, C. et al. (2006) The Marine Viromes of Four Oceanic Regions. *PLOS Biology* **4**: e368.

Avgousti, D.C., Herrmann, C., Kulej, K., Pancholi, N.J., Sekulic, N., Petrescu, J. et al. (2016) A core viral protein binds host nucleosomes to sequester immune danger signals. *Nature* **535**: 173-177.

Bell, J.J. (2008) The functional roles of marine sponges. *Estuar Coast Shelf Sci* **79**: 341-353.

Bourne, D.G., Morrow, K.M., and Webster, N.S. (2016) Insights into the coral microbiome: Underpinning the health and resilience of reef ecosystems: Biological and chemical characteristics of the coral gastric cavity. *Annu Rev Microbiol* **70**: 317-340.

Bourne, D.G., Dennis, P.G., Uthicke, S., Soo, R.M., Tyson, G.W., and Webster, N. (2013) Coral reef invertebrate microbiomes correlate with the presence of photosymbionts. *ISME J* **7**: 1452-1458.

Breitbart, M., and Rohwer, F. (2005) Here a virus, there a virus, everywhere the same virus? *Trends in Microbiology* **13**: 278-284.

Brister, J.R., Ako-Adjei, D., Bao, Y., and Blinkova, O. (2015) NCBI viral genomes resource. *Nucleic Acids Res* **43**: D571-577.

Brum, J.R., Ignacio-Espinoza, J.C., Roux, S., Doulcier, G., Acinas, S.G., Alberti, A. et al. (2015) Ocean plankton. Patterns and ecological drivers of ocean viral communities. *Science* **348**: 1261498.

Cachet, N., Genta-Jouve, G., Ivanisevic, J., Chevaldonné, P., Sinniger, F., Culioli, G. et al. (2015) Metabolomic profiling reveals deep chemical divergence between two morphotypes of the zoanthid *Parazoanthus axinellae*. *Sci Rep* **5**: 8282.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.

Ceh, J., Kilburn, M.R., Cliff, J.B., Raina, J.-B., van Keulen, M., and Bourne, D.G. (2013) Nutrient cycling in early coral life stages: *Pocillopora damicornis* larvae provide their algal symbiont (*Symbiodinium*) with nitrogen acquired from bacterial associates. *Ecol Evol* **3**: 2393-2400.

Claverie, J.M. (2009) Mimivirus and *Mimiviridae*: giant viruses with an increasing number of potential hosts, including corals and sponges. *J Invertebr Pathol* **101**: 172-180.

Colson, P., De Lamballerie, X., Yutin, N., Asgari, S., Bigot, Y., Bideshi, D.K. et al. (2013) "*Megavirales*", a proposed new order for eukaryotic nucleocytoplasmic large DNA viruses. *Arch Virol* **158**: 2517-2521.

Correa, A.M., Welsh, R.M., and Vega Thurber, R.L. (2013) Unique nucleocytoplasmic dsDNA and +ssRNA viruses are associated with the dinoflagellate endosymbionts of corals. *ISME J* **7**: 13-27.

Correa, A.M.S., Ainsworth, T.D., Rosales, S.M., Thurber, A.R., Butler, C.R., and Vega Thurber, R.L. (2016) Viral outbreak in corals associated with an *in situ* bleaching event: Atypical herpes-like viruses and a new *Megavirus* infecting *Symbiodinium*. *Front Microbiol* **7**.

Costello, E.K., Lauber, C.L., Hamady, M., Fierer, N., Gordon, J.I., and Knight, R. (2009) Bacterial Community Variation in Human Body Habitats Across Space and Time. *Science* **326**: 1694-1697.

Davy, J.E., and Patten, N.L. (2007) Morphological diversity of virus-like particles within the surface microlayer of scleractinian corals. *Aquat Microb Ecol* **47**: 37-44.

de Freitas, D.S., Coelho, M.C., Souza, M.T., Jr., Marques, A., and Ribeiro, E.B. (2007) Introduction of the anti-apoptotic baculovirus p35 gene in passion fruit induces herbicide tolerance, reduced bacterial lesions, but does not inhibits passion fruit woodiness disease progress induced by cowpea aphid-borne mosaic virus (CABMV). *Biotechnol Lett* **29**: 79-87.

Dillingham, M.S., and Kowalczykowski, S.C. (2008) RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiol Mol Biol Rev* **72**: 642-671.

Dinsdale, E., Edwards, R., Hall, D., Angly, F., Breitbart, M., Brulc, J. et al. (2008) Functional metagenomic profiling of nine biomes. *Nature*.

Doxey, A.C., Kurtz, D.A., Lynch, M.D.J., Sauder, L.A., and Neufeld, J.D. (2014) Aquatic metagenomes implicate Thaumarchaeota in global cobalamin production. *ISME J* **9**: 461.

Erives, A.J. (2015) Eukaryotic core histone diversification in light of the histone doublet and DNA topo II genes of *Marseilleviridae*. *bioRxiv*.
Fan, L., Reynolds, D., Liu, M.Y., Stark, M., Kjelleberg, S., Webster, N., and Thomas, T. (2012) Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. *Proc Natl Acad Sci U S A* **109**: E1878.
Fiore, C.L., Jarett, J., K.,, Olson, N.D., and Lesser, M.P. (2010) Nitrogen fixation and nitrogen transformations in marine symbioses. *Trends Microbiol* **18**: 455.
François, S., Filloux, D., Roumagnac, P., Bigot, D., Gayral, P., Martin, D.P. et al. (2016) Discovery of parvovirus-related sequences in an unexpected broad range of animals. *Sci Rep* **6**: 30880.
Galperin, M.Y., Makarova, K.S., Wolf, Y.I., and Koonin, E.V. (2015) Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res* **43**: D261-D269.
Gauthier, M.-E.A., Watson, J.R., and Degnan, S.M. (2016) Draft genomes shed light on the dual bacterial symbiosis that dominates the microbiome of the coral reef sponge *Amphimedon queenslandica*. *Front Mar Sci* **3**.
Gillespie, K.A., Mehta, K.P., Laimins, L.A., and Moody, C.A. (2012) Human papillomaviruses recruit cellular DNA repair and homologous recombination factors to viral replication centers. *J Virol* **86**: 9520-9526.
Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J. et al. (2013) De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nat Protoc* **8**: 10.1038/nprot.2013.1084.
Haller, S.L., Peng, C., McFadden, G., and Rothenburg, S. (2014) Poxviruses and the evolution of host range and virulence. *Infect Genet Evol* **0**: 15-40.
Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M.C. et al. (2015) eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res*.
Hughes, A.L., Irausquin, S., and Friedman, R. (2010) The evolutionary biology of poxviruses. *Infect Genet Evol* **10**: 50.
Huson, D.H., Mitra, S., Ruscheweyh, H.-J., Weber, N., and Schuster, S.C. (2011) Integrative analysis of environmental sequences using MEGAN4. *Genome Res* **21**: 1552-1560.
John, S.G., Mendez, C.B., Deng, L., Poulos, B., Kauffman, A.K., Kern, S. et al. (2011) A simple and efficient method for concentration of ocean viruses by chemical flocculation. *Environ Microbiol Rep* **3**: 195-202.
Kelly, L.W., Williams, G.J., Barott, K.L., Carlson, C.A., Dinsdale, E.A., Edwards, R.A. et al. (2014) Local genomic adaptation of coral reef-associated microbiomes to gradients of natural variability and anthropogenic stressors. *Proceedings of the National Academy of Sciences* **111**: 10227-10232.
Kimes, N.E., Van Nostrand, J.D., Weil, E., Zhou, J., and Morris, P.J. (2010) Microbial functional structure of *Montastraea faveolata*, an important Caribbean reef-building coral, differs between healthy and yellow-band diseased colonies. *Environ Microbiol* **12**: 541-556.
King, A.M., Lefkowitz, E., Adams, M.J., and Carstens, E.B. (2011) *Virus taxonomy: ninth report of the International Committee on Taxonomy of Viruses*: Elsevier.
Krabbe, M., and Carlson, K. (1991) In vivo restriction. Sequence and structure of endonuclease II-dependent cleavage sites in bacteriophage T4 DNA. *J Biol Chem* **266**: 23407-23415.
Kristensen, D.M., Kannan, L., Coleman, M.K., Wolf, Y.I., Sorokin, A., Koonin, E.V., and Mushegian, A. (2010) A low-polynomial algorithm for assembling clusters of orthologous groups from intergenomic symmetric best matches. *Bioinformatics* **26**: 1481-1487.
Krupovic, M., and Koonin, E.V. (2014) Evolution of eukaryotic single-stranded DNA viruses of the *Bidnaviridae* family from genes of four other groups of widely different viruses. *Sci Rep* **4**: 5347.
Kushmaro, A., Banin, E., Loya, Y., Stackebrandt, E., and Rosenberg, E. (2001) *Vibrio shiloi sp.* nov., the causative agent of bleaching of the coral *Oculina patagonica*. *Int J Syst Evol Microbiol* **51**: 1383-1388.

Kuss, S.K., Best, G.T., Etheredge, C.A., Pruijssers, A.J., Frierson, J.M., Hooper, L.V. et al. (2011) Intestinal microbiota promote enteric virus replication and systemic pathogenesis. *Science (New York, NY)* **334**: 249-252.
La Scola, B., Audic, S., Robert, C., Jungang, L., de Lamballerie, X., Drancourt, M. et al. (2003) A giant virus in amoebae. *Science* **299**: 2033.
Labonté, J., and Suttle, C. (2013) Metagenomic and whole-genome analysis reveals new lineages of gokushoviruses and biogeographic separation in the sea. *Front Microbiol* **4**.
Laffy, P.W., Wood Charlson, E., Turaev, D., Weynberg, K.D., Botte, E., van Oppen, M.J. et al. (2016) HoloVir: A workflow for investigating the diversity and function of viruses in invertebrate holobionts. *Front Microbiol* **7**.
Lesser, M.P., Falcón, L.I., Rodríguez-Román, A., Enríquez, S., Hoegh-Guldberg, O., and Iglesias-Prieto, R. (2007) Nitrogen fixation by symbiotic cyanobacteria provides a source of nitrogen for the scleractinian coral *Montastraea cavernosa*. *Mar Ecol Prog Ser* **346**: 143-152.
Levin, R.A., Voolstra, C.R., Weynberg, K.D., and van Oppen, M.J.H. (2017) Evidence for a role of viruses in the thermal sensitivity of coral photosymbionts. *ISME J* **11**: 808-812.
Li, J., Chen, Q., Long, L.-J., Dong, J.-D., Yang, J., and Zhang, S. (2014) Bacterial dynamics within the mucus, tissue and skeleton of the coral *Porites lutea* during different seasons. *Sci Rep* **4**: 7320.
Li, W., and Godzik, A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**: 1658-1659.
Lieberman, P.M. (2006) Chromatin regulation of virus infection. *Trends Microbiol* **14**: 132-140.
Lindell, D., Jaffe, J.D., Johnson, Z.I., Church, G.M., and Chisholm, S.W. (2005) Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* **438**: 86.
Littman, R., Willis, B.L., and Bourne, D.G. (2011) Metagenomic analysis of the coral holobiont during a natural bleaching event on the Great Barrier Reef. *Environ Microbiol Rep* **3**: 651.
Ludeman, D.A., Reidenbach, M.A., and Leys, S.P. (2016) The energetic cost of filtration by demosponges and their behavioural response to ambient currents. *J Exp Biol*.
Luter, H.M., Whalan, S., and Webster, N.S. (2010) Exploring the role of microorganisms in the disease-like syndrome affecting the sponge *Ianthella basta*. *Appl Environ Microbiol* **76**: 5736-5744.
Luter, H.M., Whalan, S., and Webster, N.S. (2012) The marine sponge *Ianthella basta* can recover from stress-induced tissue regression. *Hydrobiologia* **687**: 227-235.
Marhaver, K.L., Edwards, R.A., and Rohwer, F. (2008) Viral communities associated with healthy and bleaching corals. *Environ Microbiol* **10**: 2277-2286.
Minot, S., Bryson, A., Chehoud, C., Wu, G.D., Lewis, J.D., and Bushman, F.D. (2013) Rapid evolution of the human gut virome. *Proceedings of the National Academy of Sciences* **110**: 12450-12455.
Morrow, K.M., Moss, A.G., Chadwick, N.E., and Liles, M.R. (2012) Bacterial associates of two Caribbean coral species reveal species-specific distribution and geographic variability. *Appl Environ Microbiol* **78**: 6438-6449.
Morrow, K.M., Bourne, D.G., Humphrey, C., Botte, E.S., Laffy, P., and Zaneveld, J. (2014) Natural volcanic CO2 seeps reveal future trajectories for host-microbial associations in corals and sponges. *ISME J* **9**: 894-908.
Neary, J.M., Powell, A., Gordon, L., Milne, C., Flett, F., Wilkinson, B. et al. (2007) An asparagine oxygenase (AsnO) and a 3-hydroxyasparaginyl phosphotransferase (HasP) are involved in the biosynthesis of calcium-dependent lipopeptide antibiotics. *Microbiology* **153**: 768-776.
Nelson, M., Zhang, Y., and Van Etten, J.L. (1993) DNA methyltransferases and DNA site-specific endonucleases encoded by chlorella viruses. In *DNA Methylation: Molecular Biology and Biological Significance*. Jost, J.-P., and Saluz, H.-P. (eds). Basel: Birkhäuser Basel, pp. 186-211.
Noguchi, H., Taniguchi, T., and Itoh, T. (2008) MetaGeneAnnotator: Detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. *DNA Res* **15**: 387-396.

Patten, N.L., Harrison, P.L., and Mitchell, J.G. (2008) Prevalence of virus-like particles within a staghorn scleractinian coral (*Acropora muricata*) from the Great Barrier Reef. *Coral Reefs* **27**: 569-580.
Perron, M.-C., and Juneau, P. (2011) Effect of endocrine disrupters on photosystem II energy fluxes of green algae and cyanobacteria. *Environ Res* **111**: 520-529.
Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., and Yarza, P. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**: D590-D596.
Reyes, A., Haynes, M., Hanson, N., Angly, F.E., Heath, A.C., Rohwer, F., and Gordon, J.I. (2010) Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* **466**: 334.
Rosario, K., Breitbart, M., Harrach, B., Segales, J., Delwart, E., Biagini, P., and Varsani, A. (2017) Revisiting the taxonomy of the family *Circoviridae*: establishment of the genus *Cyclovirus* and removal of the genus *Gyrovirus*. *Arch Virol* **162**: 1447-1463.
Roux, S., Krupovic, M., Poulet, A., Debroas, D., and Enault, F. (2012) Evolution and diversity of the *Microviridae* viral family through a collection of 81 new complete genomes assembled from virome reads. *PLoS ONE* **7**: e40418.
Roux, S., Krupovic, M., Debroas, D., Forterre, P., and Enault, F. (2013) Assessment of viral community functional potential from viral metagenomes may be hampered by contamination with cellular sequences. *Open Biol* **3**.
Roux, S., Brum, J.R., Dutilh, B.E., Sunagawa, S., Duhaime, M.B., Loy, A. et al. (2016) Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* **537**: 689-693.
Ruetzler, K. (2004) Sponges on coral reefs: a community shaped by competitive cooperation. *Boll Mus Ist Biol Univ Genova* **68**: 85-148.
Salmond, G.P.C., and Fineran, P.C. (2015) A century of the phage: past, present and future. *Nat Rev Micro* **13**: 777-786.
Schmieder, R., and Edwards, R. (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**: 863-864.
Shah, N., Hülsmeier, A.J., Hochhold, N., Neidhart, M., Gay, S., and Hennet, T. (2014) Exposure to *Mimivirus* collagen promotes arthritis. *J Virol* **88**: 838-845.
Simpson, T.L. (1984) *The cell biology of sponges*: Springer-Verlag.
Soffer, N., Brandt, M.E., Correa, A.M.S., Smith, T.B., and Thurber, R.V. (2014) Potential role of viruses in white plague coral disease. *ISME J* **8**: 271-283.
Sogin, E.M., Putnam, H.M., Nelson, C.E., Anderson, P., and Gates, R.D. (2017) Correspondence of coral holobiont metabolome with symbiotic bacteria, archaea and *Symbiodinium* communities. *Environ Microbiol Rep* **9**: 310-315.
Sullivan, M.B., Coleman, M.L., Weigele, P., Rohwer, F., and Chisholm, S.W. (2005) Three *Prochlorococcus* cyanophage genomes: Signature features and ecological interpretations. *PLoS Biol* **3**: e144.
Sun, G., Xiao, J., Wang, H., Gong, C., Pan, Y., Yan, S., and Wang, Y. (2014) Efficient purification and concentration of viruses from a large body of high turbidity seawater. *MethodsX* **1**: 197-206.
Sunagawa, S., Woodley, C.M., and Medina, M. (2010) Threatened corals provide underexplored microbial habitats. *PLoS ONE* **5**: e9554.
Sunagawa, S., DeSantis, T.Z., Piceno, Y.M., Brodie, E.L., DeSalvo, M.K., and Voolstra, C.R. (2009) Bacterial diversity and White Plague Disease-associated community changes in the Caribbean coral *Montastraea faveolata*. *ISME J* **3**: 512-521.
Székely, A.J., and Breitbart, M. (2016) Single-stranded DNA phages: from early molecular biology tools to recent revolutions in environmental microbiology. *FEMS Microbiol Lett* **363**: fnw027-fnw027.
Taylor, M.W., Radax, R., Steger, D., and Wagner, M. (2007) Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. *Microbiol Mol Biol Rev* **71**: 295-347.
Taylor, M.W., Tsai, P., Simister, R.L., Deines, P., Botte, E., Ericson, G. et al. (2013) 'Sponge-specific' bacteria are widespread (but rare) in diverse marine environments. *ISME J* **7**: 438-443.

The UniProt Consortium (2015) UniProt: a hub for protein information. *Nucleic Acids Res* **43**: D204-D212.
Thomas, T., Moitinho-Silva, L., Lurgi, M., Björk, J.R., Easson, C., Astudillo-García, C. et al. (2016) Diversity, structure and convergent evolution of the global sponge microbiome. *Nat Commun* **7**: 11870.
Thurber, R.V., Haynes, M., Breitbart, M., Wegley, L., and Rohwer, F. (2009) Laboratory procedures to generate viral metagenomes. *Nat Protocols* **4**: 470-483.
Trewick, S.C., Henshaw, T.F., Hausinger, R.P., Lindahl, T., and Sedgwick, B. (2002) Oxidative demethylation by *Escherichia coli* AlkB directly reverts DNA base damage. *Nature* **419**: 174-178.
Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M.E., Borchers, C.H., Tempst, P., and Zhang, Y. (2006) Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **439**: 811-816.
Vacelet, J., and Donadey, C. (1977) Electron microscope study of the association between some sponges and bacteria. *J Exp Mar Biol Ecol* **30**: 301-314.
Vega Thurber, R., Payet, J.P., Thurber, A.R., and Correa, A.M.S. (2017) Virus-host interactions and their roles in coral reef health and disease. *Nat Rev Micro* **15**: 205-216.
Vega Thurber, R.L., Barott, K.L., Hall, D., Liu, H., Rodriguez-Mueller, B., Desnues, C. et al. (2008) Metagenomic analysis indicates that stressors induce production of herpes-like viruses in the coral *Porites compressa*. *Proc Natl Acad Sci U S A* **105**: 18413-18418.
Wang, Y., Naumann, U., Wright, S.T., and Warton, D.I. (2012) mvabund– an R package for model-based analysis of multivariate abundance data. *Methods Ecol Evol* **3**: 471-474.
Webster, N.S., and Thomas, T. (2016) The sponge hologenome. *mBio* **7**: e00135.
Wegley, L., Edwards, R., Rodriguez-Brito, B., Liu, H., and Rohwer, F. (2007) Metagenomic analysis of the microbial community associated with the coral *Porites astreoides*. *Environ Microbiol* **9**: 2707 - 2719.
Welsh, R.M., Zaneveld, J.R., Rosales, S.M., Payet, J.P., Burkepile, D.E., and Thurber, R.V. (2016) Bacterial predation in a marine host-associated microbiome. *ISME J* **10**: 1540-1544.
Weynberg, K.D., Wood-Charlson, E.M., Suttle, C.A., and van Oppen, M.J. (2014) Generating viral metagenomes from the coral holobiont. *Front Microbiol* **5**: 206.
Weynberg, K.D., Neave, M., Clode, P.L., Voolstra, C.R., Brownlee, C., Laffy, P. et al. (2017) Prevalent and persistent viral infection in cultures of the coral algal endosymbiont *Symbiodinium*. *Coral Reefs*: 1-12.
Williams, T. (2008) Natural invertebrate hosts of iridoviruses (*Iridoviridae*). *Neotrop Entomol* **37**: 615-632.
Wilson, W.H., Dale, A.L., Davy, J.E., and Davy, S.K. (2005) An enemy within? Observations of virus-like particles in reef corals. *Coral Reefs* **24**: 145-148.
Wood-Charlson, E.M., Weynberg, K.D., Suttle, C.A., Roux, S., and van Oppen, M.J. (2015) Metagenomic characterization of viral communities in corals: mining biological signal from methodological noise. *Environ Microbiol*.
Wu, L., Qiu, Z., Zhou, Y., Du, Y., Liu, C., Ye, J., and Hu, X. (2016) Physiological effects of the herbicide glyphosate on the cyanobacterium *Microcystis aeruginosa*. *Aquat Toxicol* **178**: 72-79.
Zhang, J., Kobert, K., Flouri, T., and Stamatakis, A. (2014) PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**: 614-620.
Zhang, Y., Nelson, M., Nietfeldt, J., Xia, Y., Burbank, D., Ropp, S., and Van Etten, J.L. (1998) Chlorella virus NY-2A encodes at least 12 DNA endonuclease/methyltransferase genes. *Virology* **240**: 366-375.
Zhong, C., Edlund, A., Yang, Y., McLean, J.S., and Yooseph, S. (2016) Metagenome and metatranscriptome analyses using protein family profiles. *PLoS Comput Biol* **12**: e1004991.

Figure Legends

Figure 1. Community composition and sequence similarity of holobiont associated virome communities. NMDS plot based on the Bray Curtis similarity of genus-level taxonomic assignment of predicted genes from virome communities.

Figure 2. Composition of viruses across holobiont species and reef environments. Taxonomic conservation and OTU estimates were calculated using HMMgraspx to analyse individual reads from virome data sets and by comparison to HMMER models generated from VOGDB viral orthologous groups pertaining to major helicase, polymerase, terminase and major capsid proteins. Reads were normalised across samples and total abundance was calculated based on the number of reads that were assigned to each orthologous group. OTU estimates were generated based on the number of reads recruited to selected VOGs and adjusted according to the number of helicase, polymerase, terminase and major capsid proteins that are typically found in each viral family. Coverage is depicted by bubble size and the number of OTUs by bubble colour.

Figure 3. Top 50 Swissprot keywords assigned to holobiont viromes. Swissprot functional keyword assignment to predicted gene data was adjusted to normalise for keyword composition within the Swissprot database. Keyword composition was adjusted to account for the coverage of the source contig within the virome community. The top 50 most abundant keywords across all datasets were identified and the corresponding COG functional category assignment of the source genes were assigned according to best EggNOG v4.5 BLAST+ match. COG functional categories: (F) Nucleotide metabolism and transport; (L) Replication and repair; (M) Cell wall/membrane/envelope biogenesis; (S) Function unknown.

Figure 4. Viral functions that were significantly different between reef host species. MVabund was used to perform univariate tests on Swissprot keyword abundance data from all samples, identifying key drivers of functional differences between host species. Swissprot keyword assignments were normalised to keyword composition within the Swissprot database and adjusted to account for the coverage of the source contig within the virome community.

Figure 5. Viral functions that were significantly different between reef environment (coral, sponge, seawater). MVabund was used to perform univariate tests on Swissprot keyword abundance data from all samples, identifying key drivers of functional differences between reef environments. Swissprot keyword assignments were normalised to keyword composition within the Swissprot database and adjusted to account for the coverage of the source contig within the virome community.

Figure 6. Analysis workflow of coral, sponge and seawater virome datasets.

Table Legends

Table 1. Host species resemblance based on SIMPER analysis showing average similarity in viral community composition between hosts. Permutations were based on a Bray-Curtis similarity matrix generated from square root transformed data. Samples with significant P-values from pairwise tests (<0.05) are highlighted in bold.

	Sponges				**Corals**		**Seawater**	
	A. queenslandica	*X. testudinaria*	*R. odorabile*	*I. basta*	*P. acuta*	*P. lutea*	GBR seawater	PNG seawater
A. queenslandica	81.192							
X. testudinaria	74.008	78.868						
R. odorabile	**66.042**	**65.852**	86.378					
I. basta	69.152	**67.611**	**60.123**	78.297				
P. acuta	62.698	**58.234**	**51.701**	**59.254**	66.183			
P. lutea	**74.114**	**69.491**	**58.996**	**67.952**	**59.218**	76.423		
GBR seawater	75.631	**66.572**	**64.853**	**68.162**	**63.751**	**69.833**	82.018	
PNG seawater	**72.126**	**73.167**	**60.148**	**71.726**	**58.032**	**74.709**	**67.289**	88.286

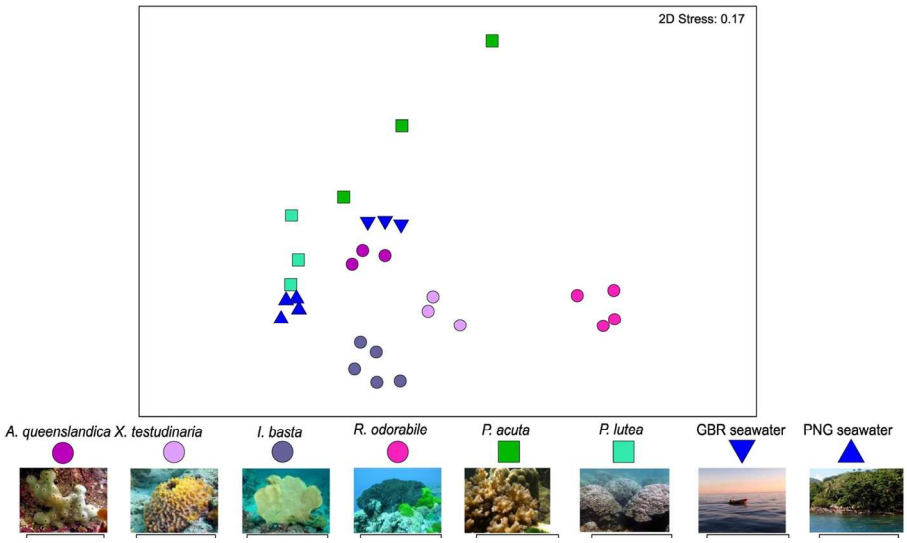

Figure 1. Community composition and sequence similarity of holobiont associated virome communities. NMDS plot based on the Bray Curtis similarity of genus-level taxonomic assignment of predicted genes from virome communities.

136x73mm (300 x 300 DPI)

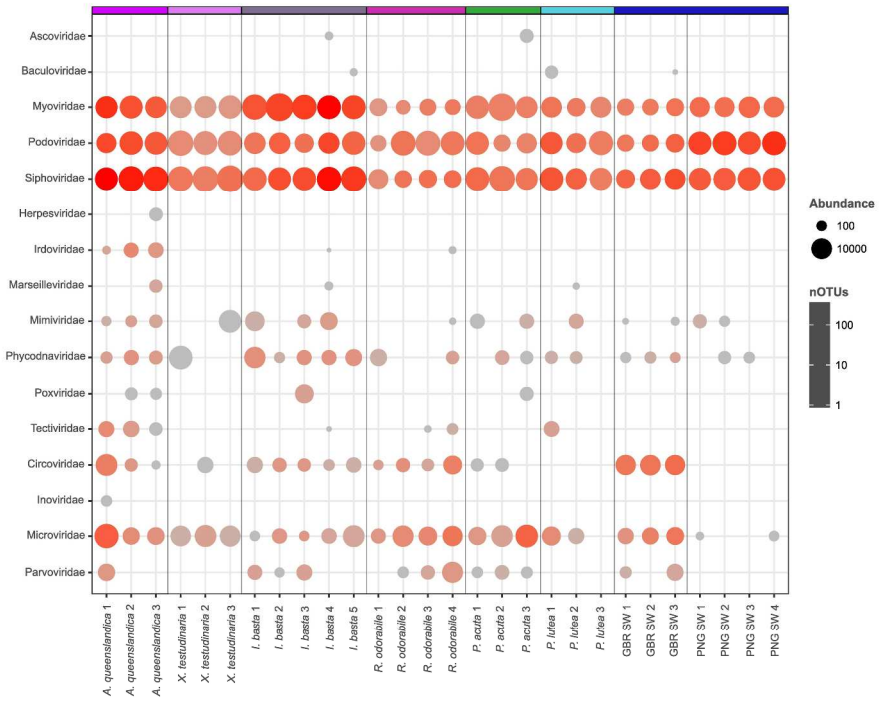

Figure 2. Composition of viruses across holobiont species and reef environments. Taxonomic conservation and OTU estimates were calculated using HMMgraspx to analyse individual reads from virome data sets and by comparison to HMMER models generated from VOGDB viral orthologous groups pertaining to major helicase, polymerase, terminase and major capsid proteins. Reads were normalised across samples and total abundance was calculated based on the number of reads that were assigned to each orthologous group. OTU estimates were generated based on the number of reads recruited to selected VOGs and adjusted according to the number of helicase, polymerase, terminase and major capsid proteins that are typically found in each viral family. Coverage is depicted by bubble size and the number of OTUs by bubble colour.

279x361mm (300 x 300 DPI)

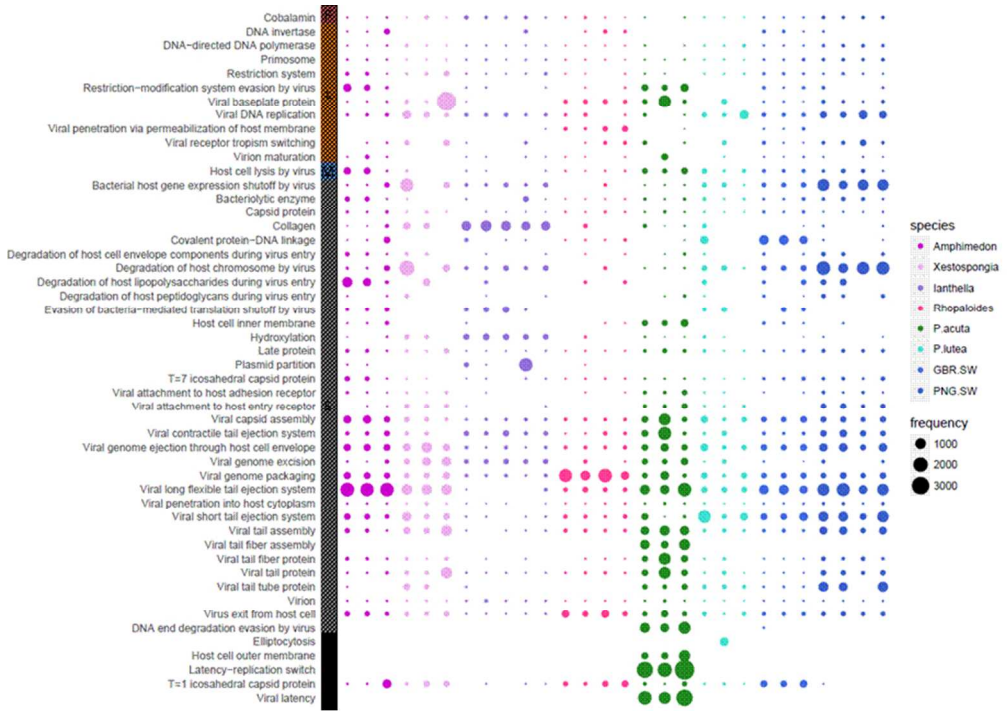

Figure 3. Top 50 Swissprot keywords assigned to holobiont viromes. Swissprot functional keyword assignment to predicted gene data was adjusted to normalise for keyword composition within the Swissprot database. Keyword composition was adjusted to account for the coverage of the source contig within the virome community. The top 50 most abundant keywords across all datasets were identified and the corresponding COG functional category assignment of the source genes were assigned according to best EggNOG v4.5 BLAST+ match. COG functional categories: (F) Nucleotide metabolism and transport; (L) Replication and repair; (M) Cell wall/membrane/envelope biogenesis; (S) Function unknown.

322x227mm (72 x 72 DPI)

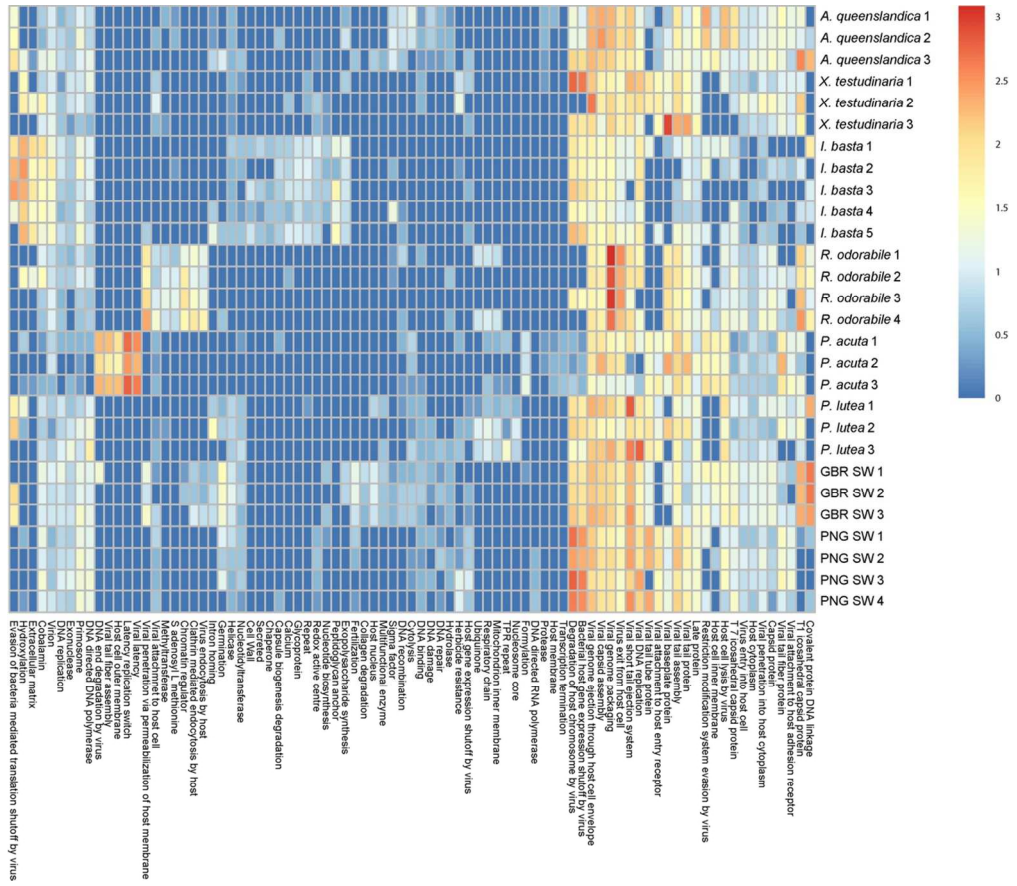

Figure 4. Viral functions that were significantly different between reef host species. MVabund was used to perform univariate tests on Swissprot keyword abundance data from all samples, identifying key drivers of functional differences between host species. Swissprot keyword assignments were normalised to keyword composition within the Swissprot database and adjusted to account for the coverage of the source contig within the virome community.

118x103mm (300 x 300 DPI)

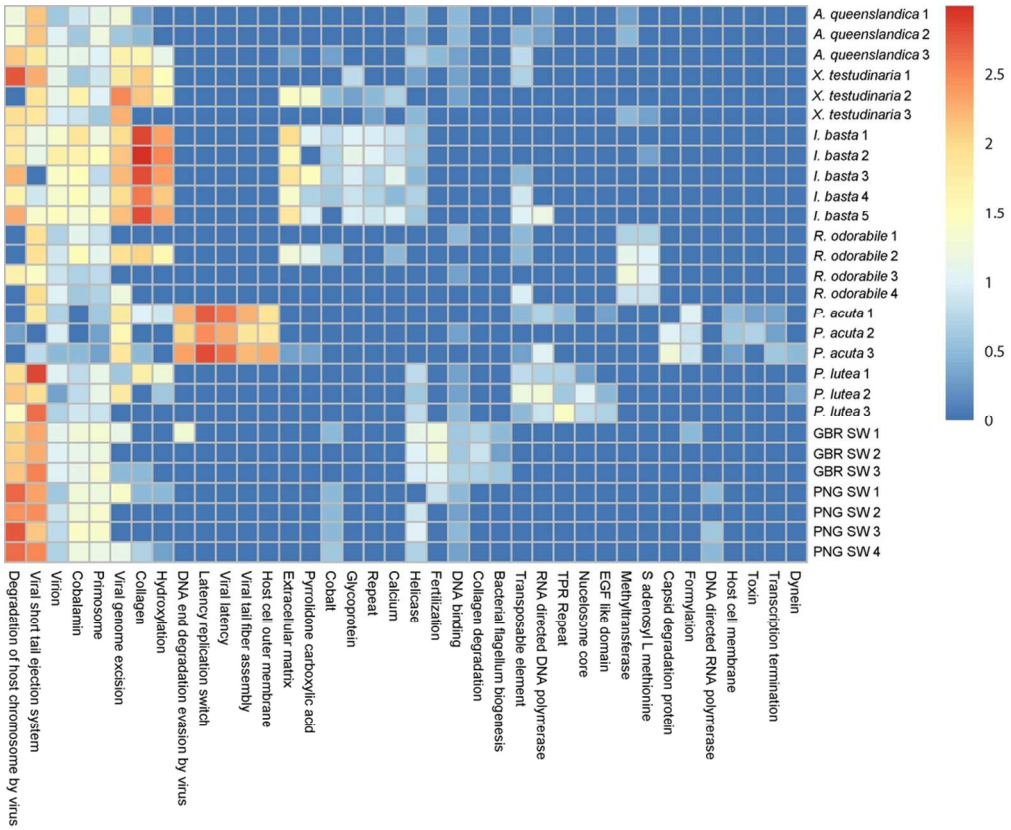

Figure 5. Viral functions that were significantly different between reef environment (coral, sponge, seawater). MVabund was used to perform univariate tests on Swissprot keyword abundance data from all samples, identifying key drivers of functional differences between reef environments. Swissprot keyword assignments were normalised to keyword composition within the Swissprot database and adjusted to account for the coverage of the source contig within the virome community.

104x86mm (300 x 300 DPI)

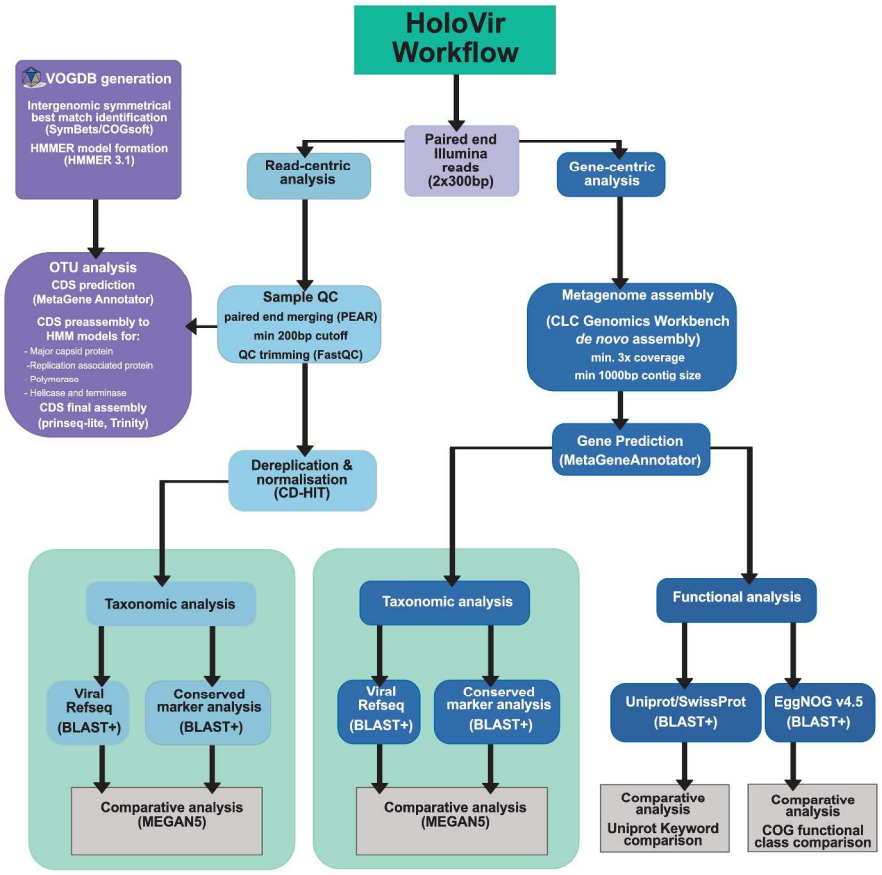

Figure 6. Analysis workflow of coral sponge and seawater virome datasets.

282x266mm (300 x 300 DPI)

Author/s:
Laffy, PW; Wood-Charlson, EM; Turaev, D; Jutz, S; Pascelli, C; Botte, ES; Bell, SC; Peirce, TE; Weynberg, KD; van Oppen, MJH; Rattei, T; Webster, NS

Title:
Reef invertebrate viromics: diversity, host specificity and functional capacity

Date:
2018-06-01

Citation:
Laffy, P. W., Wood-Charlson, E. M., Turaev, D., Jutz, S., Pascelli, C., Botte, E. S., Bell, S. C., Peirce, T. E., Weynberg, K. D., van Oppen, M. J. H., Rattei, T. & Webster, N. S. (2018). Reef invertebrate viromics: diversity, host specificity and functional capacity. ENVIRONMENTAL MICROBIOLOGY, 20 (6), pp.2125-2141. https://doi.org/10.1111/1462-2920.14110.