Taxonomic, functional and expression analysis of viral communities associated with marine sponges

Mary Nguyen ¹, Bernd Wemheuer ¹, Patrick Laffy ², Nicole Webster ²,³, Torsten Thomas Corresp. ¹

¹ Centre for Marine Science and Innovation & School of Biological & Earth and Environmental Sciences, University of New South Wales, Sydney, Australia
² Australian Institute of Marine Science, Townsville, Australia
³ Australian Centre for Ecogenomics, University of Queensland, Brisbane, Australia

Corresponding Author: Torsten Thomas
Email address: t.thomas@unsw.edu.au

Viruses play an essential role in shaping the structure and function of ecological communities. Marine sponges have the capacity to filter large volumes of “virus-laden” seawater through their bodies and host dense communities of microbial symbionts, which are likely accessible to viral infection. However, despite the potential of sponges and their symbionts to act as viral reservoirs, little is known about the sponge-associated virome. Here we address this knowledge gap by analysing metagenomic and (meta-)transcriptomic datasets from several sponge species to determine what viruses are present and elucidate their predicted and expressed functionality. Sponges were found to carry diverse, abundant and active bacteriophages as well as eukaryotic viruses belonging to the Megavirales and Phycodnaviridae. These viruses contain and express auxiliary metabolic genes (AMGs) for photosynthesis and vitamin synthesis as well as for the production of antimicrobials and the defence against toxins. These viral AMGs can therefore contribute to the metabolic capacities of their hosts and also potentially enhance the survival of infected cells. This suggest that viruses may play a key role in regulating the abundance and activities of members of the sponge holobiont.
Taxonomic, functional and expression analysis of viral communities associated with marine sponges.

Mary THD Nguyen¹, Bernd Wemheuer¹, Patrick Laffy², Nicole Webster²,³, Torsten Thomas¹

¹ Centre for Marine Science and Innovation & School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, Australia

² Australian Institute of Marine Science, Townsville, Australia

³ Australian Centre for Ecogenomics, University of Queensland, Brisbane, Australia

Corresponding authors: Torsten Thomas, t.thomas@unsw.edu.au

Running title: Function of the sponge virome
Viruses play an essential role in shaping the structure and function of ecological communities. Marine sponges have the capacity to filter large volumes of “virus-laden” seawater through their bodies and host dense communities of microbial symbionts, which are likely accessible to viral infection. However, despite the potential of sponges and their symbionts to act as viral reservoirs, little is known about the sponge-associated virome. Here we address this knowledge gap by analysing metagenomic and (meta-)transcriptomic datasets from several sponge species to determine what viruses are present and elucidate their predicted and expressed functionality. Sponges were found to carry diverse, abundant and active bacteriophages as well as eukaryotic viruses belonging to the Megavirales and Phycodnaviridae. These viruses contain and express auxiliary metabolic genes (AMGs) for photosynthesis and vitamin synthesis as well as for the production of antimicrobials and the defence against toxins. These viral AMGs can therefore contribute to the metabolic capacities of their hosts and also potentially enhance the survival of infected cells. This suggest that viruses may play a key role in regulating the abundance and activities of members of the sponge holobiont.
Viruses, and in particular bacteriophages, are thought to be the most abundant biological entities in the ocean. With an estimated $10^7$-$10^8$ virus-like particles (VLPs) per millilitre of seawater, they outnumber their bacterial hosts up to 100-fold (Breitbart et al. 2018; Fuhrman 1999; Knowles et al. 2016; Parikka et al. 2017; Weinbauer 2004; Wigington et al. 2016; Wommack & Colwell 2000). In the marine environment, viruses are considered to play an important role in: 1) global biogeochemical cycles through lysis of microorganisms (Middelboe & Lyck 2002; Weinbauer et al. 2011), 2) microbial community structure and diversity through the ‘kill the winner’ process (Angly et al. 2006; Winter et al. 2010) and 3) the evolution of microorganisms through viral-mediated genetic exchange (Breitbart 2012; Fuhrman & Schwalbach 2003; Riemann & Middelboe 2002; Rodriguez-Brito et al. 2010; Thurber et al. 2017).

The influence of viruses on microbial community structure and evolution are of particular interest for filter-feeding reef invertebrates, such as sponges (phylum Porifera), due to their capacity to move large volumes of “virus-laden” seawater through their bodies and their hosting of dense communities of microbial symbionts, which are likely accessible to viral infection (Reiswig 1971; Thomas et al. 2010). Despite the potential of sponges and their symbionts to act as viral reservoirs, the investigation of sponge-associated viruses is still in its infancy. Early electron microscopy studies by Vacelet and Gallissian (1978) detected VLPs within Verongia cavernicola tissue that was devoid of choanocytes chambers, while Claverie et al. (2009) later predicted that the phagocytes of Petrobiona massiliana may be undergoing infection by a giant virus, possibly related to the Mimiviridae. More recently, Pascelli et al. (2018) detected over 50 VLP morphotypes within the cells, the extracellular matrix (mesohyl) and the associated microorganisms of 15 different sponge species.

Metagenomic sequencing has also recently been applied to taxonomically describe and functionally characterise the communities of free viruses associated with sponge species (Jahn et al. 2019; Laffy et al. 2018). This work found diverse viral assemblages that exhibited significant host species-specificity and revealed several auxiliary metabolic genes (AMGs) and viral pathogenesis pathways that were differentially enriched across sponge hosts (Laffy et al. 2018). AMGs encode proteins for metabolic functions that are thought to increase the fitness of the infected host e.g. by allowing for a better utilisation of nutrients or energy sources. These viral communities were further dominated by bacteriophages (e.g. Myoviridae, Podoviridae and Siphoviridae), but had considerable variability in the relative abundance of viruses targeting eukaryotic cells (such as Mimiviridae and Phycodnaviridae). The role of viruses in host health and environmental stress responses has also been recently started to be explored and demonstrated shifts in viral community profiles (Butina et al. 2019; Laffy et al. 2018). Furthermore evidence for genetic exchange through sponge-associated viruses has been given by the discovery of
eukaryotic-like genes encoding for ankyrin proteins in bacteriophages that may provide infected bacterial hosts with an increased capacity to avoid eukaryotic phagocytosis (Jahn et al. 2019). Several studies have also reported indirect evidence for the presence of viruses in sponges, specifically through the presence of antiviral defence mechanisms, such as clustered regularly interspaced palindromic repeats (CRISPR), that have been identified within genome and metagenome sequences of sponge-associated microbes (Fan et al. 2012; Horn et al. 2016; Karimi et al. 2017; Thomas et al. 2010). And finally, viral genomes have also been found within the microbial metagenome of a hydrothermal vent sponge in the Okinawa trough (Zhou et al. 2019).

While most metagenomic studies have focused on viral communities that were physically separated from their microbial or eukaryotic hosts, investigating the viral diversity that is directly associated with the cellular fraction may enhance our understanding of the role viruses play in the sponge holobiont. This study therefore aimed to investigate the abundance, diversity and functional gene repertoire of viruses associated with metagenomic datasets derived from the microbial cells of six sponge species, including a comparative analysis with viruses associated with microorganisms from the surrounding seawater. In addition, no previous study has investigated what viruses and functions are actively expressed in sponges and we therefore used here a (meta-) transcriptomic approach for the species *C. concentrica*, *Scopalina sp.*, and *T. anhelans* to fill this knowledge gap.

**Materials and Methods**

**Metagenomic, transcriptomic and metatranscriptomic datasets**

To retrieve viral sequences associated with microbial cells, metagenomic datasets from six marine sponges (*Cymbastela concentrica*, *Tedania anhelans*, *Scopalina sp.* from the coast of Sydney, Australia and *Cymbastela coralliophila*, *Rhopaloeides odorabile* and *Stylissa sp.* 445 from the Great Barrier Reef, Australia; July-September 2009) and corresponding seawater metagenomes previously published by Fan et al. (2012) were analysed (Figure S1). Microbial cells in that study were obtained by homogenisation of sponge tissue followed by a series of centrifugation and filtration steps (3 µm cut-off), which removed sponge cells and likely removed free (i.e. not cell-bound) VLPs (Castro-Mejia et al. 2015; Fan et al. 2012). Seawater microorganisms were collected by filtration on a 0.2 µm membrane. DNA from the microbial cells of each sample type were extracted, shotgun sequenced using Roche pyrosequencing technology and reads as well as assembled contigs were annotated (further details can be found in Fan et al. 2012). Taxonomic and functional analyses showed high intra-species consistency and clear inter-species differences for the bacterial and archaeal communities of these sponge species that were also all distinct from those in seawater (Fan et al. 2012). Expressed viral genes were analysed in the transcriptomic and
metatranscriptomic datasets from the sponges *C. concentrica*, *Scopalina sp.* and *T. anhelans* (sampled in July 2016), which were recently published by Diez-Vives et al. (2017) (Figure S1). Transcriptomic and metatranscriptomic sequencing data were generated in that study from total RNA extractions of snap-frozen whole sponge tissue. Total RNA was separated into poly-adenylated (polyA) RNA and non-polyA RNA, with the latter being additionally depleted of rRNA, to represent enrichment of mRNA transcript of eukaryotes and mRNA of prokaryotes, respectively. These two RNA fraction were shotgun-sequenced with the Illumina sequencing platform followed by a *de novo* assembly (further details can be found in Diez-Vives et al. 2017) and are here after referred to as “transcriptomic” and “metatranscriptomic” data, respectively. The datasets combined therefore represent non-rRNA transcripts from microbial and sponge cells as well as all other constituents of the sponge holobiont (Diez-Vives et al. 2017).

**Identification of viral sequences**

The VirSorter program (Roux et al. 2015) was used to identify viral sequences within reads and contigs of the cellular metagenomes and (meta-)transcriptomes mentioned above. VirSorter uses an in-built ‘viromes’ reference database, in which 826 proteins clusters and single genes were identified as “viral hallmark genes”. Virsorter calculates the following parameters: 1) presence of viral hallmark genes (Koonin et al. 2006; Roux et al. 2014), 2) enrichment in virus-like genes 3) depletion in PFAM-affiliated genes, 4) enrichment in uncharacterized genes (i.e. predicted genes with no hits in either PFAM or the viral reference database), 5) enrichment in short genes (genes with a size within the 10% of the shortest genes of the genome or contig), and 6) depletion in strand switching (i.e. change of coding strand between two consecutive genes). Contigs were subsequently classified based on these parameters into three categories (i) category 1 (“most confident” viral predictions) for contigs fulfilling criteria 1 and 2; (ii) category 2 (“likely” viral predictions) for contigs fulfilling either criteria 1 or 2 and at least one of the other criteria; and (iii) category 3 (“possible” viral predictions) for contigs not fulfilling criteria 1 or 2, but at least two of the other criteria with at least one significance score greater than 4 (for more details see (Roux et al. 2015)). In this study, only contigs that were assigned to category 1 and 2 were used for further analyses. The viral sequence identified in the cellular metagenomes and (meta-)transcriptomes are available as the supplementary data file “Supplementary_data_viral_contigs_transcripts.fna” with the ID description given in the supplementary information.

Viral sequences from the cellular metagenomes were normalised to read coverage as well as the estimated number of prokaryotic genomes (Fan et al. 2012). The number of prokaryotic genomes per sample was inferred using the average abundance of 82 single-copy genes, which were present in ≥ 90% of all complete archaeal and bacterial genomes in the NCBI RefSeq database (archaea = 261, bacteria = 7315; July 2017) and were found as single-copies in ≥95% of all genomes. The PFAM and TIGRFAM HMM profiles (version 31.0 and 14.0, respectively) of the proteins encoded by these single-copy gene
were used for homolog identification in the metagenomes using HMMsearch v3.1b2 (minimum score of 40 and maximum E-value of \(10^{-5}\)). Viral sequences in the (meta-) transcriptomes were normalised using transcripts per million (TPM) counts (Diez-Vives et al. 2017).

**Taxonomic and functional annotation**

Taxonomic classification of viral sequences was conducted using MetaVir2 (Roux et al. 2014) using the viral Refseq protein database from NCBI with BLASTp and an E-value threshold of \(10^{-3}\) as well as the PFAM database v27 with HMMScan and a score threshold of 30. Taxonomic classification was based on the lowest common ancestor (LCA) affiliation of each contig as implemented in MetaVir2 (Roux et al., 2014). The LCA affiliation considers multiple hits on a single sequence (up to five, if available) and affiliation is made at the highest common taxonomic level from the selected genes against the RefSeqVirus database.

For functional annotation of sequences, ORFs were predicted using Prodigal v.2.6.2 (Hyatt et al. 2010). Predicted protein sequences were functionally classified using Diamond searches (Buchfink et al. 2015) against the COG (2014 release) (Galperin et al. 2014) and KEGG (October 2017 release) databases (Kanehisa et al. 2002) with an E value of \(<10^{-10}\). In addition, predicted proteins were compared to the PFAM v31 (Finn et al. 2015) and TIGRFAM v15 (Haft et al. 2003) databases using HMMsearch with an E-value of \(<10^{-10}\). Functional profiles for each sample are represented by relative abundance counts for each functional category. Proteins with more than one assigned functional category or domain contributed to each function/functional domain equally. Profiles of the KEGG and COG pathways for each sample were calculated from the abundance of all affiliated specific functions. Abundance of functional categories belonging to more than one pathway were equally divided between the pathways.

Heatmaps, community diversity and statistical analysis were conducted in R (R core Team 2014) using the vegan package (Oksanen et al. 2013).

**Results**

**Viral diversity associated with microbial cells of sponges**

Viral sequences made up a small percentage (0-0.47%) of the total sequences within each metagenomic dataset (Table 1). In total, 451 viral sequences were identified using VirSorter’s most stringent criteria (Roux et al. 2015), of which 344 could be taxonomically classified to species level using MetaVir’s LCA
strategy (Roux et al. 2014). Cellular metagenomes from seawater contained on average 16 viral sequences per prokaryotic genome and the sponges Scopalina sp., C. coralliophila, Stylissa sp. 445, C. concentrica, T. anhelans and R. odorabile contained an average sum of 3.6, 2.7, 2.5, 1.2, 1.1 and 0.68 contigs per prokaryotic genome, respectively.

Viromes from all samples were dominated by members of the dsDNA order Caudovirales. Members of the family Myoviridae had the highest relative abundances in seawater followed by Siphoviridae and Podoviridae. The relative abundances of these three Caudovirales families differed among each sponge species (Figures 1A & S2). Alpha diversity as measured by Shannon’s indices on the species and family level showed seawater to contain the highest viral diversity, followed by Scopalina sp., C. coralliophila, T. anhelans, R. odorabile, C. concentrica and Stylissa sp. 445 (Table S1). High variability among individual replicates and between different host species was observed in viral assemblages at both the species and family levels (Figures 1, S2-S4). Significant differences in the taxonomic structure of the viral communities across all sponge species was seen at the species and family levels (PERMANOVA, Df=5, F.Model=1.6254, R²=0.42489 P=0.001 and F.Model=1.918, R²=0.46575, P=0.014, respectively), however no support was found for significant pairwise differences between sponge species (Table S2). Additionally, PERMANOVA did not reveal any significant differences between taxonomic structure of sponge and seawater viral assemblages at the family or species level (PERMANOVA, Df=2; F.Model= 1.1182, R²=0.11626, P= 0.243 and F.Model=1.4267, R²=0.14373, P=0.111, respectively).

Viral communities in the meta-transcriptome and transcriptome

The percentage of viral contigs in the metatranscriptomes of C. concentrica and Scopalina sp. were 0.003% and 0.002%, respectively, and no viral contigs could be identified in the metatranscriptome of T. anhelans. Low percentages of viral contigs were also identified in the host transcriptome of C. concentrica and T. anhelans (0.002% and 0.0005%, respectively) (Table 2) and no viral contigs could be identified in the transcriptome of Scopalina sp.. Only DNA viruses were identified in the (meta-)transcriptomic data, with 12 viral sequences identified in the metatranscriptome of C. concentrica and two in Scopalina sp., while five viral sequences were identified in the transcriptome of C. concentrica and a single sequence was identified within the transcriptome of T. anhelans (Table 2).

The viral metatranscriptome of C. concentrica was primarily comprised of Caudovirales belonging to the families Myoviridae and Siphoviridae, but also contained representatives from the eukaryotic virus family Phycodnaviridae (Figure 1B). In C. concentrica, a relative of Ostreococcus lucimarinus virus 7 (family Phycodnaviridae) had an average expression greater than three-fold the average total TPM of the metatranscriptome (Figure 1B). Metatranscriptomes from Scopalina sp. contained two eukaryotic viruses, with closest similarity to Pandoravirus dulcis and Ostreococcus lucimarinus virus 7, which were
expressed approximately equal to the average TPM of the metatranscriptome. (Figure 2). Expression of Phycodnaviridae genes was consistent with the presence of Phycodnaviridae contigs in the Scopalina sp. metagenome (Figure 1A).

Four sequences within the C. concentrica transcriptome were taxonomically assigned to bacteriophages of Cronobacter, Klebsiella, Bacillus and Synechococcus. The Cronobacter phage sequence in the C. concentrica transcriptomes was expressed at approximately twice the average TPM, and sequences belonging to this phage taxon were also observed in the metagenomic and metatranscriptomic datasets of this sponge. Sequences assigned to Klebsiella, Bacillus and Synechococcus phages were not found in corresponding metatranscriptome or metagenome datasets. The T. anhelans transcriptome also contained a viral sequences assigned to the family Phycodnaviridae (Figure 1B).

Functional gene analysis of viral contigs

Functional gene annotation showed that 16%, 20%, 21.4% and 13% of all predicted genes within viral sequence could be mapped to the KEGG, COG, PFAM and TIGRFAM databases, respectively. Consistent with taxonomic analysis, functional profiles from metagenome-derived viral sequences showed high variability between individual replicates and host species (Figure 2, Figure S5-8). Functional categories of the metagenomic sequences assigned to viruses were distinct from the whole (i.e. viral plus cellular) metagenomes (Figure 2A, B). Not surprisingly, COG annotation associated with the categories “mobility: prophages, transposon” and “replication, recombination and repair” were enriched in the viral sequences compared to the whole metagenomes (Figure 2A). Abundant viral functions detected in the PFAM, KEGG, COG and TIGRFAM annotations comprised “Mu-like prophage proteins”, “DNA/RNA polymerase” and “DNA primase and helicases” proteins (Figures S5-8). The KEGG pathway “Pantothenate and CoA biosynthesis” was found in the viral metagenomes of C. concentrica and the “purine and pyrimidine metabolism” pathway was found in Stylissa sp. 445 (Figure 2B). Several other genes with potential auxiliary functions for bacterial or eukaryotic hosts were also detected within the viral sequences, including pathways for biosynthesis of antibiotics, genes for aegerolysin and bacterial toxin-antitoxin systems (ydaT and HicA) (Figure 2B, C and Figure S5). Genes involved in “energy production and conversion”, and more specifically photosynthesis functions, were observed in viral sequences from the C. concentrica, R. odorabile and Stylissa sp. 445 metagenomes (Figure 2C & S5).

Functional analysis based on (meta-)transcriptomics data showed the relative abundance of expressed genes associated with the COG category “Post-translational modification, protein turnover, chaperones” (Figure 3A). Viral sequences from the C. concentrica metatranscriptomes contained expressed genes assigned to the COG categories “amino acid transport and metabolism” and “transcription” and viral sequences from the Scopalina sp. metatranscriptomes included genes associated with the COG category “replication, recombination and repair” (no KEGG pathways were identified for viral sequences derived...
from the metatranscriptomes of *Scopalina sp* (Figure 3A and B). Genes associated with the KEGG pathways “thiamine metabolism”, “sulfur relay systems” and “antibiotic biosynthesis” were all identified in the viral sequences within the metatranscriptomes of *C. concentrica* (Figure 3B).

**Discussion**

**Viruses associated with microbial cells in sponges**

Here we present the first dedicated analysis of viral sequences in a cell-separated metagenomic dataset of sponges, representing viruses attached to prokaryotic cells, virions present inside cells and/or prophages. Most bacteriophage sequences were predicted to be dsDNA viruses, although ssDNA viruses (such as *Inoviridae*) were also detected (Figure 1A). Caudovirales were the dominant bacteriophage order, comprising the viral families, *Myoviridae, Siphoviridae* and *Podoviridae* (Maniloff & Ackermann 1998). Caudovirales have previously been reported as the dominant component of free-living metaviromes derived from sponges (Laffy et al. 2016; Laffy et al. 2018; Weynberg et al. 2017; Williamson et al. 2012) as well as viromes derived from seawater and soil (Clokie et al. 2011; Zablocki et al. 2014).

Across the viral sequences retrieved in this study, numerous matches to the Megavirales were also identified, including matches to the *Marseilleviridae, Mimiviridae*, and *Pithoviridae* (Figure 1A). Previous analyses of sponges also detected the ubiquitous presence of different families of Megavirales, with the authors suggesting that these viruses may infect amoeba-like phagocytic sponge cells (known as amoebocytes) or some other components of the sponge holobiont (Laffy et al. 2019; Laffy et al. 2016; Laffy et al. 2018). The presence of sequences assigned to Megavirales is also consistent with microscopic observations, which show “parasitic infections” of sponge amoebocytes that resembled micrographs of *Acanthamoeba* infected with *Mimivirus* (Claverie et al. 2009).

Viral sequences associated with the microbial cell fraction of the sponges investigated here showed high variability within and across sponge species, both in terms of community structure and composition. This contrasts other recent studies of sponge-associated viromes performed by Laffey et al. (2018 and 2019), which showed high intra-species similarity in viral assemblages. Possible explanations for this discrepancy are that Laffy et al. investigated amplified DNA extracted from the free viral fraction, while our study focused on unamplified DNA from the microbial cell fraction, or that samples in the two studies were taken at different times and locations, or that the scarcity of viral DNA in the microbial cell metagenomes (Bzhalava et al. 2018) limited sequence coverage, which led to variable assembly success of viral contigs in each sample (Smits et al. 2015). Despite this discrepancy, detection of specific viral sequences within
microbial metagenomic and (meta-) transcriptomic datasets indicate that the viral taxa and functions
identified here are playing core roles within the sponge holobionts studied here, as they likely constitute
dominant or active viruses.

Active viral consortia in sponges

Active viruses in *C. concentrica* and *Scopalina sp.* were generally observed at high expression levels,
equal to or above the average TPM of any gene transcript in the sample. For example, *Phycodnaviridae*
and *Pandoraviridae* were observed in both sponges' (meta-) transcriptomes at 2-3 fold higher TPM values
than the average TPM of other genes in *C. concentrica* and at comparable values to the average TPM in
*Scopalina sp.*

*Phycodnaviridae* belong to the superfamily of nucleocytoplasmic large DNA viruses (NCLDV), which infect
marine and freshwater algae (Koonin & Yutin 2010) and are also prevalent in the free-virus fraction of the
marine sponges *Amphimedon queenslandica* and *Ianthella basta* (Laffy et al. 2018). *C. concentrica* has
previously been reported to contain a high abundance of diatoms (Taylor et al. 2004), and these diatoms
could potentially play host to these *Phycodnaviridae*, although so far no diatom-infecting NCLDV has been
described in the literature. Sponges may also filter and concentrate phototrophic organisms that are host
to *Phycodnaviridae* from the seawater (Laffy et al. 2018). The observation that the phycodnaviruses are
closely related to *Ostreococcus lucimarinus* virus (Figure 1B) suggests that *C. concentrica* may contain a
microalgal family similar to *O. lucimarinus*, which has been found in coastal and mesotrophic systems in
the Atlantic and Pacific oceans and the Mediterranean Sea (Derelle et al. 2015).

Expressed bacteriophage sequences were only detected in the *C. concentrica* transcriptomes and
metatranscriptomes (Figure 1B) and these were taxonomically assigned to *Cronobacter, Burkholderia,*
*Aeromonas, Enterobacteria* and *Shigella* phages. Some of these bacteriophage taxa were also found in
the metagenomic data (Figure 1A). While sequences related to these bacteriophages are clearly
expressed and hence suggest ‘active’ viruses, sequences corresponding to their bacterial hosts have so
far not been found in *C. concentrica*, despite extensive metagenomic and 16S rRNA amplicon based
analyses (Esteves et al. 2016; Fan et al. 2012). A potential explanation for this observation is that these
viral taxa have a broader or different host-range than their taxonomic or phylogenetic affiliation implies.
However this requires further investigation.
Approximately 80% of ORFs annotated from viral sequences could not be assigned as being homologous to proteins with known functions, which is consistent with a high percentage of ‘unknown’ proteins (generally >70%) commonly observed in viral metagenome studies (Adriaenssens et al. 2015; Hurwitz & Sullivan 2013; Laffy et al. 2018; Weynberg et al. 2017; Williamson et al. 2012; Zablocki et al. 2014). Generally, genes associated with ‘replication, recombination and repair’, ‘protein metabolism’ and ‘carbohydrate metabolism’ were abundant in all viral metagenomes and (meta-) transcriptomes, consistent with enrichment of these functional categories in virome studies from different environments, including sponges (Laffy et al. 2018), corals (Weynberg et al. 2017), seawater (Williamson et al. 2012) and soil (Adriaenssens et al. 2015; Zablocki et al. 2014).

While the broad functional potential of viral communities in sponges has been previously reported, our analysis of metagenomic, transcriptomic and metatranscriptomic datasets provided a unique opportunity to identify dominant and actively expressed viral functions. Analysis of viral metagenomic sequences revealed several AMGs with potential benefits to bacterial or eukaryotic hosts, such as genes associated with the biosynthesis of antibiotics, aegerolysin, bacterial toxin-antitoxin systems (ydaT and HicA) and photosystem II (PSII) (Figure 2A-C). Analysis of the viral transcripts from the (meta-) transcriptomic data further revealed the expression of genes associated with antibiotic biosynthesis, photosynthesis and thiamine metabolism (Figure 3).

A range of genes assigned to KEGG pathways for antibiotic synthesis and resistance were expressed in viruses of C. concentrica, including streptomycin and vancomycin biosynthesis, acarbose and validamycin biosynthesis, general biosynthesis of antibiotics and cationic antimicrobial peptide (CAMP) resistance (Figure 3). Viruses encoding and expressing these pathways may enhance their bacterial host’s fitness through inhibition of competitor or defence against the many antibiotics that have been reported to be produced within the sponge holobiont (Faulkner 1978; Hentschel et al. 2001; Kelman et al. 2009; Kim et al. 2006; Laport et al. 2009; Sipkema et al. 2005; Torres et al. 2002). Similarly, Laffy et al. (2018) reported the enrichment of herbicide-resistance genes, which were assigned to Synechococcus phages, in the sponge Xestospongia testudinaria. Herbicides are highly effective in controlling cyanobacterial populations and it was postulated that the Synechococcus phages confer herbicide resistance to their cyanobacterial hosts and enhanced their host’s survival (Laffy et al. 2018).

A high relative abundance of aegerolysin genes was detected in viruses found in the Scopalina sp. metagenomes (average abundance of 7.8 copies per prokaryotic genome; Figure 2C). Proteins of the aegerolysin family are widely distributed in fungi and bacteria, and to a lesser extent in plants, protozoa...
and insects (Berne et al. 2009; Butala et al. 2017). The functions of aegerolysins are not well understood, but they are thought to have broad-range biological properties, including antitumor, antiproliferation and antibacterial activities (Berne et al. 2009; Butala et al. 2017). Interestingly, the ascovirus Trichoplusia ni 2c, an obligate viral pathogen of Pseudoplusia includens larvae and other insects from the family Noctuidae, was reported to encode a hypothetical aegerolysin-like protein (Wang et al. 2006), suggesting that aegerolysin genes may provide benefits to the host, such as an antibacterial activity. Chemical extracts from sponges and their associated microorganisms are reported to contain antitumor, antiproliferation and antibacterial activities (Laport et al. 2009; Lee et al. 2001; Monks et al. 2002; Thakur et al. 2003; Yung et al. 2011), and it may be therefore possible that viruses may contribute to these features via horizontal gene transfer.

A putative bacterial toxin gene ydaT was found in the virome of Scopalina sp. at an average of 6.7 copies per prokaryotic genome (Figure 2C) and a HicA toxin was actively expressed in the viral metatranscriptome of C. concentrica (Figure S10). Toxin-antitoxin systems are widely distributed in bacteria and are associated with the formation of antibiotic-tolerant (persister) cells (Butt et al. 2014). YdaT belongs to the type II toxin-antitoxin systems (Sevin & Barloy-Hubler 2007), where ydaS expresses a toxin and YdaT constitutes the antitoxin (Yamaguchi & Inouye 2011). It was recently reported that YdaS and YdaT may both act as toxins by inhibiting the regular cell division of E. coli, which may provide resistance against β-lactamase antibiotics (Bindal et al. 2017). Interestingly, the HicA toxins from Burkholderia pseudomallei have also been reported to arrest cell growth and increase the number of persister cells tolerant to ciprofloxacin and ceftazidime (Butt et al. 2014). Many sponges contain high amounts of antibacterial compounds (see above) and the HicA/YdaT-type proteins may perhaps therefore play a role in providing resistance against them.

Genes for the PSII-associated proteins PsbA and PsbD were expressed in a viral contigs from Stylissa sp. 445 (Figure S6), which were taxonomically assigned to Synechococcus phages (Fan et al. 2012). In addition, genes for the PS II reaction centre were identified from viral sequences identified in the C. concentrica and R. odorabile metagenomes (Figure 2C). Interestingly, psbA and psbD have also been recently found in coral viromes (Laffy et al. 2018; Weynberg et al. 2017) and in cyanophages infecting free-living Synechococcus and Prochlorococcus (Lindell et al. 2004; Mann et al. 2003; Weigele et al. 2007). Viral PSII genes are thought to supplement photosynthesis of their host and boost energy production, presumably for virion production (Clokie et al. 2006; Lindell et al. 2005). Genes encoding the carboxysome-shell carbonic anhydrase were also detected in the virome of Stylissa sp. 445 (Figure 2C). Carbonic anhydrase is responsible for catalysing the conversion of carbonic acid to CO₂ and is utilised by cyanobacteria and marine phytoplankton to support photosynthetic carbon fixation (Badger et al. 2005; Reinfelder 2010).
Thiamine (vitamin B1) functions as a cofactor for many essential metabolic pathways and is only produced by bacteria, fungi, and plants (Jurgenson et al. 2009). Genes associated with thiamine metabolism were expressed in the viral metatranscriptome of *C. concentrica* (Table S3). Specifically, the expression of genes for the tyrosine decarboxylase and many cysteine desulfurase enzymes were detected (Table S3). Cysteine desulfurase is involved in the initial stages of sulfur trafficking, contributing to the addition of sulfur into the thiazole ring of thiamine (Begley et al. 1999; Mihara & Esaki 2002) (Figure S12). To our knowledge, this is the first time that viral-associated genes involved in thiamine metabolism have been observed, and this could potentially confer growth advantages to the host cells that cannot produce vitamin B1 (such as sponge cells). Genes associated with other vitamin synthesis pathways, such as cobalamin (vitamin B12), have also been reported in the virome of other sponges (Laffy et al. 2018) and in myoviruses that infect *Prochlorococcus* (Sullivan et al. 2005).

**Conclusion**

Recent targeted virome analyses have discovered that marine viruses can encode AMGs previously thought to be restricted to host genomes (Enav et al. 2014; Hurwitz et al. 2013; Laffy et al. 2018; Weynberg et al. 2017). Here we demonstrate that viral AMGs can also be identified through analysis of sponge-associated metagenomic, transcriptomic and metatranscriptomic datasets, providing a unique perspective on abundant and active viruses within the sponge holobiont. Viruses contain and express AMGs for photosynthesis and vitamin synthesis, as well as for the production of and defence against toxins and antimicrobials. These viral AMGs therefore could not only contribute to the metabolic capacities of their hosts, but also potentially enhance survival of infected cells. This model suggests that viruses may play a key role in regulating the abundance and activities of members of the sponge holobiont. However, further experimental work will be required to confirm or define the role that these AMGs have in the physiology and ecology of their host.

**Acknowledgement:**

This work was supported by funds provided by the Betty and Gordon Moore Foundation and the Australian Research Council to TT and an Australian Research Council Future Fellowship (FT120100480) to NW.
**Figure Legends**

**Figure 1:** Viral communities identified within sponges and seawater metagenomes (A) and (meta-)transcriptomes (B). Samples were clustered based on Bray-Curtis dissimilarities of taxonomic profiles (tree scale indicates dissimilarity percentages). (A) Values are normalised to viral genomes per prokaryotic genome transformed with fourth root. Numbers before taxonomic assignments indicate the number of viral species found within each taxon. Taxa highlighted in bold indicate unclassified taxa above genus level and corresponding numbers in brackets indicate the number of contigs. Alphabet A, B and C indicate sample replicates for sponges and number 01, 02 and 08 indicate replicates for seawater samples, corresponding the nomenclature in Fan et al. 2012. Numbers 1, 2, and 3 indicate individual sponge metatranscriptome replicates. (B) For transcriptome data, three replicates were pooled together before sequencing. Values are natural logs of normalised data transcripts per million (TPM), coloured markers on the scale represent average TPM of all transcripts in the corresponding samples coded by colours, orange: *Scopalina sp.* meta-transcriptome, red: *T. anhelans* transcriptome, purple: *C. concentrica* meta-transcriptome and navy: *C. concentrica* transcriptome.

**Figure 2:** Functional composition of metagenome-derived viral communities. Functional enrichment based on COG annotations (A), KEGG pathway annotations (B), and PFAM and TIGRFAM database (C). For all analyses, annotations were assigned to assembled predicted genes normalised to proportion of gene copies per prokaryotic genome and transformed with cube root. (A) and (B) contain annotations to the corresponding whole microbial metagenomes (viral + bacterial) for comparative purposes. Samples are clustered with Bray-Curtis dissimilarity using the ‘average’ method, tree scale is based on dissimilarity percentages. Letters A, B and C indicate individual sponge replicates and numbers 01, 02 and 08 indicate seawater replicates.

**Figure 3:** Functional gene analysis of sponge viral and whole (viral and prokaryotic) meta-transcriptomes and transcriptomes. Functional enrichment based on COG annotations (A) and KEGG pathway annotations (B) of assembled predicted genes were normalised to transcripts per million (TPM) and transformed with square root. Samples are clustered with Bray-Curtis dissimilarity using the ‘average’ method, tree scale is based on dissimilarity percentages. Numbers 1, 2 and 3 indicate individual sponge replicates. For transcriptomic data, the three replicates were pooled together.
References

Adriaenssens EM, Van Zyl L, De Maayer P, Rubagotti E, Rybicki E, Tuffin M, and Cowan DA. 2015. Metagenomic analysis of the viral community in Namib Desert hypoliths. *Environmental microbiology* 17:480-495.

Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, Carlson C, Chan AM, Haynes M, Kelley S, and Liu H. 2006. The marine viromes of four oceanic regions. *PLoS biology* 4:e368.

Badger MR, Price GD, Long BM, and Wooldger FJ. 2005. The environmental plasticity and ecological genomics of the cyanobacterial CO2 concentrating mechanism. *Journal of experimental botany* 57:249-265.

Begley TP, Xi J, Kinsland C, Taylor S, and McLafferty F. 1999. The enzymology of sulfur activation during thiamin and biotin biosynthesis. *Current opinion in chemical biology* 3:623-629.

Berne S, Lah L, and Sepčić K. 2009. Aegerolysins: structure, function, and putative biological role. *Protein Science* 18:694-706.

Bindal G, Krishnamurthi R, Seshasayee ASN, and Rath D. 2017. CRISPR-Cas-mediated gene silencing reveals RacR to be a negative regulator of YdaS and YdaT toxins in *Escherichia coli* K-12. *mSphere* 2:e00483-00417.

Breitbart M. 2012. Marine viruses: truth or dare. *Annual Reviews in Marine Sciences* 4:425-448.

10.1146/annurev-marine-120709-142805

Breitbart M, Bonnain C, Malki K, and Sawaya NA. 2018. Phage puppet masters of the marine microbial realm. *Nature Microbiology* 1:754-766.

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

Buchfink B, Xie C, and Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nature methods* 12:59.

Butala M, Novak M, Krašiček M, Skočaj M, Veranič P, Maček P, and Sepčić K. 2017. Aegerolysins: lipid-binding proteins with versatile functions. *Seminars in Cell & Developmental Biology* Elsevier.

Butina T, Bukin YS, Khanaei I, Kravtsova L, Maikova O, Tupikin A, Kabilov M, and Belikov S. 2019. Metagenomic analysis of viral communities in diseased Baikal sponge *Lubomirskia baikalensis*. *Limnology and Freshwater Biology*:155-162.

Butt A, Higman VA, Williams C, Crump MP, Hemsley CM, Harner M, and Titball RW. 2014. The HicA toxin from *Burkholderia pseudomallei* has a role in persister cell formation. *Biochemical Journal* 459:333-344.

Bzhalava Z, Tampuu A, Bala P, Vicente R, and Dillner J. 2018. Machine Learning for detection of viral sequences in human metagenomic datasets. *BMC bioinformatics* 19:336.

Castro-Meija JL, Muhammed MK, Kot W, Neve H, Franz CM, Hansen LH, Vogensen FK, and Nielsen DS. 2015. Optimizing protocols for extraction of bacteriophages prior to metagenomic analyses of phage communities in the human gut. *Microbiology* 3:64.

Claverie JM, Grzelka R, Lartigue A, Bernadac A, Nitsche S, Vacelet J, Ogata H, and Abergel C. 2009. Mimivirus and Mimiviridae: giant viruses with an increasing number of potential hosts, including corals and sponges. *J Invertebr Pathol* 101:172-180. 10.1016/j.jip.2009.03.011

Clokie MR, Millard AD, Letarov AV, and Heaphy S. 2011. *Phages in nature*. *Bacteriophage* 1:31-45.

Clokie MR, Shan J, Bailey S, Jia Y, Krisch HM, West S, and Mann NH. 2006. Transcription of a "photosynthetic"T4-like phage during infection of a marine cyanobacterium. *Environmental microbiology* 8:827-835.

Derelle E, Monier A, Cooke R, Worden AZ, Grimsley NH, and Moreau H. 2015. Diversity of viruses infecting the green microalga *Ostreococcus lucimarinus*. *Journal of virology* 89:5812-5821.

Diez-Vives C, Moitinho-Silva L, Nielsen S, Reynolds D, and Thomas T. 2017. Expression of eukaryotic-like protein in the microbiome of sponges. *Molecular ecology* 26:1432-1451.

Enav H, Mandel-Gutfreund Y, and Béjà O. 2014. Comparative metagenomic analyses reveal viral-induced shifts of host metabolism towards nucleotide biosynthesis. *Microbiome* 2:9.

Esteves AIS, Amer N, Nguyen M, and Thomas T. 2016. Sample processing impacts the viability and cultivability of the sponge microbiome. *Frontiers in Microbiology* 7. 10.3389/fmicb.2016.00499

Fan L, Reynolds D, Liu M, Stark M, Kjelleberg S, Webster NS, and Thomas T. 2012. Functional equivalence and evolutionary convergence in complex communities of microbial sponge
symbionts. Proceedings of the National Academy of Sciences U S A 109:E1878-1887. 10.1073/
pnas.1203287109

Faulkner D. 1978. Antibiotics from sponges. Topics in antibiotic chemistry 2:20-29.
Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, and
Sangrador-Vegas A. 2015. The Pfam protein families database: towards a more sustainable
future. Nucleic acids research 44:D279-D285.

Fuhrman J, and Schwabach M. 2003. Viral influence on aquatic bacterial communities. The Biological
Bulletin 204:192-195.

Fuhrman JA. 1999. Marine viruses and their biogeochemical and ecological effects. Nature 399:541-548.
Galperin MY, Makarova KS, Wolf YI, and Koonin EV. 2014. Expanded microbial genome coverage and
improved protein family annotation in the COG database. Nucleic acids research 43:D261-D269.

Haft DH, Selengut JD, and White O. 2003. The TIGRFAMs database of protein families. Nucleic acids
research 31:371-373.

Hentschel U, Schmid M, Wagner M, Fieseler L, Gernert C, and Hacker J. 2001. Isolation and phylogenetic
analysis of bacteria with antimicrobial activities from the Mediterranean sponges Aplysina
aerophoba and Aplysina cavernicola. FEMS microbiology ecology 35:305-312.

Horn H, Slaby BM, Jahn MT, Bayer K, Moitinho-Silva L, Förster F, Abdelmohsen UR, and Hentschel U.
2016. An enrichment of CRISPR and other defense-related features in marine sponge-associated
microbial metagenomes. Frontiers in Microbiology 7.

Hurwitz BL, Hallam SJ, and Sullivan MB. 2013. Metabolic reprogramming by viruses in the sunlit and dark
ocean. Genome biology 14:R123.

Hurwitz BL, and Sullivan MB. 2013. The Pacific Ocean Virome (POV): a marine viral metagenomic dataset
and associated protein clusters for viral quantitative ecological analysis. PloS one 8:e57355.

Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, and Hauser LJ. 2010. Prodigal: prokaryotic gene
recognition and translation initiation site identification. BMC bioinformatics 11:119.

Jahn MT, Arkhipova K, Markert SM, Stigloher C, Lachnit T, Pita L, Kupczok A, Ribes M, Stengel ST, and
Rosenstiel P. 2019. A phage protein aids bacterial symbionts in eukaryote immune evasion. Cell
Host & Microbe 26:542-550, e545.

Jurgenson CT, Begley TP, and Ealick SE. 2009. The structural and biochemical foundations of thiamin
biosynthesis. Annual review of biochemistry 78:569-603.

Kanehisa M, Goto S, Kawashima S, and Nakaya A. 2002. The KEGG databases at GenomeNet. Nucleic
acids research 30:42-46.

Karimi E, Ramos M, Gonçalves J, Xavier JR, Reis MP, and Costa RC. 2017. Comparative metagenomics
reveals the distinctive adaptive features of the Spongia officinalis endosymbiotic consortium.
Frontiers in Microbiology 8:2499.

Kelman D, Kashman Y, Hill RT, Rosenberg E, and Loya Y. 2009. Chemical warfare in the sea: The search
for antibiotics from Red Sea corals and sponges. Pure and Applied Chemistry 81:1113-1121.

Kim TK, Hewavitharana AK, Shaw PN, and Fuerst JA. 2006. Discovery of a new source of rifamycin
antibiotics in marine sponge actinobacteria by phylogenetic prediction. Appl Environ Microbiol
72:2118-2125.

Knowles B, Silveira C, Bailey B, Barott K, Cantu V, Cobi n-G-Hemes A, Coutinho F, Dinsdale E, Felts B, and
Furby K. 2016. Lytic to temperate switching of viral communities. Nature 531:466-470.

Koonin EV, Senkevich TG, and Dolja VV. 2006. Discovery of a new source of rifamycin
antibiotics in marine sponge actinobacteria by phylogenetic prediction. Appl Environ Microbiol
72:2118-2125.

Koonin EV, and Yutin N. 2010. Origin and evolution of eukaryotic large nucleo-cytoplasmic DNA viruses.
Intervirology 53. 10.1159/000312913

Laffy PW, Botte ES, Wood-Charlson EM, Weynberg KD, Rattei T, and Webster NS. 2019. Thermal stress
modifies the marine sponge virome. Environmental microbiology reports 11:690-698.

Laffy PW, Wood Charlson E, Turaev D, Weynberg KD, Botte E, van Oppen MJ, Webster N, and Rattei T.
2016. HoloVir: A workflow for investigating the diversity and function of viruses in invertebrate
holobionts. Frontiers in Microbiology 7:822.

Laffy PW, Wood Charlson EM, Turaev D, Jutz S, Pascelli C, Botte ES, Bell SC, Peirce TE, Weynberg
KD, and van Oppen MJ. 2018. Reef invertebrate viromics: diversity, host specificity and functional
capacity. Environmental microbiology.

Laport M, Santos O, and Muricy G. 2009. Marine sponges: potential sources of new antimicrobial drugs.
Current pharmaceutical biotechnology 10:86-105.

Lee YK, Lee J-H, and Lee HK. 2001. Microbial symbioses in marine sponges. Journal of Microbiology
39:254-264.
Thomas T, Rusch D, DeMaere MZ, Yung PY, Lewis M, Halpern A, Heidelberg KB, Egan S, Steinberg PD, and Edwards R. 2010. Viral and microbial community dynamics in four aquatic environments. The ISME journal 4:739-751.

Roux S, Enault F, Hurwitz BL, and Sullivan MB. 2015. VirSorter: mining viral signal from microbial genomics data. PeerJ 3:e985.

Roux S, Tournayre J, Mahul A, Debroas D, and Enault F. 2014. Metavir 2: new tools for viral metagenome comparison and assembled virome analysis. BMC bioinformatics 15:76.

Sevin EW, and Barloy-Hubler F. 2007. RASTA-Bacteria: a web-based tool for identifying toxin-antitoxin loci in prokaryotes. Genome biology 8:R155.

Sipkema D, Franssen MC, Osinga R, Tramper J, and Wijffels RH. 2005. Marine sponges as pharmacy. Marine Biotechnology 7:142-162.

Smits SL, Bodewes R, Ruiz-González A, Baumgärtner W, Koopmans MP, Osterhaus AD, and Schürch AC. 2015. Recovering full-length viral genomes from metagenomes. Frontiers in Microbiology 6:1069.

Sullivan MB, Coleman ML, Weigele P, Rohwer F, and Chisholm SW. 2005. Three Prochlorococcus cyanophages genomes: signature features and ecological interpretations. PLoS biology 3:e144.

Taylor MW, Schupp PJ, Dahlöf I, Kjelleberg S, and Steinberg PD. 2004. Host specificity in marine sponge-associated bacteria, and potential implications for marine microbial diversity. Environmental microbiology 6:121-130.

Thakur NL, Hentschel U, Krasko A, Pabel CT, Anil AC, and Müller WE. 2003. Antibacterial activity of the sponge Suberites domuncula and its primmorphs: potential basis for epibacterial chemical defense. Aquatic Microbial Ecology 31:77-83.

Thomas T, Rusch D, DeMaere MZ, Yung PY, Lewis M, Halpern A, Heidelberg KB, Egan S, Steinberg PD, and Kjelleberg S. 2010. Functional genomic signatures of sponge bacteria reveal unique and shared features of symbiosis. The ISME journal 4:1557-1567. 10.1038/ismej.2010.74

Thurber RV, Payet JP, Thurber AR, and Correa AM. 2017. Virus–host interactions and their roles in coral reef health and disease. Nature Reviews Microbiology 15:205.

Torres YR, Berlinck RG, Nascimento GG, Fortier SC, Pessoa C, and de Moraes MO. 2002. Antibacterial activity against resistant bacteria and cytotoxicity of four alkaloid toxins isolated from the marine sponge Arenosclera brasiliensis. Toxicon 40:885-891.
Vacelet J. and Gallissian M-F. 1978. Virus-like particles in cells of the sponge *Verongia cavernicola* (Demospongiae, Dictyoceratida) and accompanying tissues changes. *J Invertebr Pathol* 31:246-254.

Wang L, Xue J, Seaborn CP, Arif BM, and Cheng X-W. 2006. Sequence and organization of the *Trichoplusia ni* ascovirus 2c (Ascoviridae) genome. *Virology* 354:167-177.

Weigele PR, Pope WH, Pedulla ML, Houtz JM, Smith AL, Conway JF, King J, Hatfull GF, Lawrence JG, and Hendrix RW. 2007. Genomic and structural analysis of Syn9, a cyanophage infecting marine *Prochlorococcus* and *Synechococcus*. *Environmental microbiology* 9:1675-1695.

Weinbauer MG. 2004. Ecology of prokaryotic viruses. *FEMS microbiology reviews* 28:127-181.

Weinbauer MG, Bonilla-Findji O, Chan AM, Dolan JR, Short SM, Šimek K, Wilhelm SW, and Suttle CA. 2011. *Synechococcus* growth in the ocean may depend on the lysis of heterotrophic bacteria. *Journal of plankton research* 33:1465-1476.

Weynberg KD, Laffy PW, Wood-Charlson EM, Turaev D, Rattei T, Webster NS, and van Oppen MJ. 2017. Coral-associated viral communities show high levels of diversity and host auxiliary functions. *PeerJ* 5:e4054.

Yamaguchi Y, and Inouye M. 2011. Regulation of growth and death in Escherichia coli by toxin–antitoxin systems. *Nature Reviews Microbiology* 9:779-790.

Yung PY, Burke C, Lewis M, Kjelleberg S, and Thomas T. 2011. Novel antibacterial proteins from the microbial communities associated with the sponge *Cymbastela concentrica* and the green alga *Ulva australis*. *Appl Environ Microbiol* 77:1512-1515.

Zablocki O, van Zyl L, Adriaenssens EM, Rubagotti E, Tuffin M, Cary SC, and Cowan D. 2014. High-level diversity of tailed phages, eukaryote-associated viruses, and virophage-like elements in the metaviromes of antarctic soils. *Appl Environ Microbiol* 80:6888-6897.

Zhou K, Zhang R, Sun J, Zhang W, Tian R-M, Chen C, Kawagucci S, and Xu Y. 2019. Potential interactions between Clade SUP05 Sulfur-Oxidizing Bacteria and Phages in Hydrothermal Vent Sponges. *Appl Environ Microbiol* 85.
Figure 1

Figure 1: Viral communities identified within sponges and seawater metagenomes (A) and (meta-) transcriptomes (B).

Samples were clustered based on Bray-Curtis dissimilarities of taxonomic profiles (tree scale indicates dissimilarity percentages). (A) Values are normalised to viral genomes per prokaryotic genome transformed with fourth root. Numbers before taxonomic assignments indicate the number of viral species found within each taxon. Taxa highlighted in bold indicate unclassified taxa above genus level and corresponding numbers in brackets indicate the number of contigs. Alphabet A, B and C indicate sample replicates for sponges and number 01, 02 and 08 indicate replicates for seawater samples, corresponding the nomenclature in Fan et al. 2012. Numbers 1, 2, and 3 indicate individual sponge metatranscriptome replicates. (B) For transcriptome data, three replicates were pooled together before sequencing. Values are natural logs of normalised data transcripts per million (TPM), coloured markers on the scale represent average TPM of all transcripts in the corresponding samples coded by colours, orange: Scopalina sp. meta-transcriptome, red: T. anhelans transcriptome, purple: C. concentrica meta-transcriptome and navy: C. concentrica transcriptome.
Figure 2: Functional composition of metagenome-derived viral communities.

Functional enrichment based on COG annotations (A), KEGG pathway annotations (B), and PFAM and TIGRFAM database (C). For all analyses, annotations were assigned to assembled predicted genes normalised to proportion of gene copies per prokaryotic genome and transformed with cube root. (A) and (B) contain annotations to the corresponding whole microbial metagenomes (viral + bacterial) for comparative purposes. Samples are clustered with Bray-Curtis dissimilarity using the ‘average’ method, tree scale is based on dissimilarity percentages. Letters A, B and C indicate individual sponge replicates and numbers 01, 02 and 08 indicate seawater replicates.
Manuscript to be reviewed
Figure 3: Functional gene analysis of sponge viral and whole (viral and prokaryotic) meta-transcriptomes and transcriptomes.

Functional enrichment based on COG annotations (A) and KEGG pathway annotations (B) of assembled predicted genes were normalised to transcripts per million (TPM) and transformed with square root. Samples are clustered with Bray-Curtis dissimilarity using the ‘average’ method, tree scale is based on dissimilarity percentages. Numbers 1, 2 and 3 indicate individual sponge replicates. For transcriptomic data, the three replicates were pooled together.
**Table 2**: Summary of the number of viral sequences found in the meta/transcriptome of *C. concentrica*, *Scopalina sp.* and *T. anhelans*.  

| Viral Sequence | C. concentrica | Scopalina sp. | T. anhelans |
|---------------|----------------|---------------|-------------|
| Number        |                |               |             |
**Table 2:** Summary of the number of viral sequences found in the meta/transcriptome of *C. concentrica*, *Scopalina sp.* and *T. anhelans*.

| Sample                          | No. of total sequences | No. of viral sequences | % of viral sequences in metagenome | Average length of viral sequences (bp) |
|---------------------------------|------------------------|------------------------|------------------------------------|----------------------------------------|
| Meta-transcriptome *C. concentrica* | 383648                 | 12                     | 0.003                              | 2953                                   |
| Meta-transcriptome *Scopalina sp.* | 94173                  | 2                      | 0.002                              | 3293                                   |
| Transcriptome *C. concentrica*   | 184053                 | 4                      | 0.002                              | 2614                                   |
| Transcriptome *T. anhelans*      | 129529                 | 1                      | 0.0005                             | 3313                                   |
Table 2 (on next page)

Table 1: Summary of the number of viral sequences found in each sample metagenome. A, B and C indicate sample replicates for sponges and 01, 02 and 08 indicate seawater replicates (labelling is based on Fan et al. 2012).
Table 1: Summary of the number of viral sequences found in each sample metagenome. A, B and C indicate sample replicates for sponges and 01, 02 and 08 indicate seawater replicates (labelling is based on Fan et al. 2012).

| Sample               | No. of total sequences | No. of viral sequences | % of viral sequences in metagenome | Average length of viral sequences (bp) |
|----------------------|------------------------|------------------------|-----------------------------------|---------------------------------------|
| Seawater 01          | 18797                  | 88                     | 0.468                             | 644                                   |
| Seawater 02          | 24581                  | 109                    | 0.443                             | 957                                   |
| Seawater 08          | 42686                  | 11                     | 0.026                             | 504                                   |
| C. coralliophila A   | 62543                  | 3                      | 0.005                             | 22687                                 |
| C. coralliophila B   | 45264                  | 46                     | 0.102                             | 3623                                  |
| C. coralliophila C   | 58846                  | 17                     | 0.029                             | 1252                                  |
| R. odorabile A       | 76275                  | 14                     | 0.018                             | 2600                                  |
| R. odorabile B       | 31373                  | 9                      | 0.029                             | 1169                                  |
| R. odorabile C       | 34669                  | 3                      | 0.009                             | 983                                   |
| C. concentrica A     | 27990                  | 6                      | 0.021                             | 4403                                  |
| C. concentrica B     | 26734                  | 8                      | 0.030                             | 6531                                  |
| C. concentrica C     | 68642                  | 13                     | 0.019                             | 4704                                  |
| Scopalina sp. A      | 25395                  | 21                     | 0.083                             | 3190                                  |
| Scopalina sp. B      | 30982                  | 21                     | 0.068                             | 3633                                  |
| Scopalina sp. C      | 28414                  | 27                     | 0.095                             | 2437                                  |
| T. anhelans A        | 15972                  | 11                     | 0.069                             | 3524                                  |
| T. anhelans B        | 15949                  | 10                     | 0.063                             | 3353                                  |
| T. anhelans C        | 11789                  | 7                      | 0.059                             | 1910                                  |
| Stylissa sp. 445 A   | 32276                  | 25                     | 0.077                             | 1756                                  |
| Stylissa sp. 445 B   | 21243                  | 2                      | 0.009                             | 1012                                  |
| Stylissa sp. 445 C   | 24377                  | 0                      | 0.0                                | 0                                     |