Variation in virome diversity in wild populations of *Penaeus monodon* (Fabricius 1798) with emphasis on pathogenic viruses

Fredmoore L. Orosco^{1,2} · Arturo O. Lluisma¹

Received: 15 February 2017 / Accepted: 20 June 2017 / Published online: 1 July 2017
© Indian Virological Society 2017

Abstract Marine animals typically harbor a community of viruses, a number of which are known to cause diseases. In shrimp aquaculture, viral pathogens are the principal causes of major economic losses. However, the composition of the viral load of shrimps in wild population is poorly known. In this study, we explored the viral diversity in the microbiome of wild *Penaeus monodon* collected from six sites in the Philippines, with a view to detecting pathogenic forms. We employed a metagenomic approach via particle-associated nucleic acid isolation, sequence-independent single primer amplification, and pyrosequencing. Virome analysis of shrimp samples from different sites revealed distinct virome profiles, and hence significant differences in diversity, among the various sites based on number of OTUs, Shannon–Weaver Index, and Inverse Simpson Index. Sequences of key shrimp pathogens were detected such as the white spot syndrome virus (WSSV), and *Penaeus stylirostris* densovirus (*PstDV*). However, the patterns of distribution of the pathogenic viruses varied; whereas WSSV was found only in three out of six sites and *PstDV* were found in all but one site. The results also revealed shrimp-associated viruses that have not yet been observed in *P. monodon* such as avian virus-like, insect virus-like, plankton virus-like and bacteriophage-like sequences. Despite the diverse array of viruses

detected in the study, a large proportion remains unidentified (i.e., similarity to sequences in the database was lower than the threshold required for definitive identification), and therefore could represent unexplored virus sequences and viral genomes in the environment.

Keywords Metagenomics · Shrimp pathogens · DNase-SISPA · *Penaeus monodon* · Virome

Introduction

A number of marine viruses are known to be pathogenic to *Penaeus monodon*. To date, white spot syndrome virus (WSSV) remains as the most widespread and devastating pathogen in the shrimp aquaculture industry. It has spread to almost all countries where shrimps are cultivated and remains a major threat because of its wide host range, genetic variation and changes in virulence [29]. Moreover, the *Penaeus stylirostris* densovirus (*PstDV*) has been reported to cause severe mortalities up to 90%; it has been a significant problem in the culture of several penaeid species in the Americas, Oceania and Asia [29]. Although a number of management practices have been developed, these viruses remained a significant problem in the aquaculture industry.

Our knowledge of viral diversity and viral communities (viromes) in different environmental samples has been limited by available methodology [23]. The key limitation in viral diversity studies is the lack of universal genetic markers for viruses, in contrast to bacterial diversity studies in which the 16S rRNA gene has proven to be a useful marker [18]. Recently, high throughput sequencing has been leveraged to enable a metagenomics approach to studies of uncultured viral assemblages in a process termed

✉ Arturo O. Lluisma
aolluisma@up.edu.ph

¹ Marine Genomics and Molecular Genetics Laboratory, Marine Science Institute, University of the Philippines – Diliman, 1101 Quezon City, Philippines

² Institute of Biology, University of the Philippines – Diliman, 1101 Quezon City, Philippines

as “viral metagenomics” or “viromics”. This method provides a holistic look at viral diversity within a given sample, completely bypassing the need for culturing [31]. Moreover, this approach has dramatically expanded our knowledge in understanding the diversity of viromes in a wide range of environments such as soil, sea, potable water, ballast water, activated sludge, coral, and hot springs [23, 26].

However, a key challenge in obtaining DNA for viral metagenomics studies is the fact that whole metagenomes are dominated by host DNA resulting in minimal coverage of viral DNA [3]. Thus, we employed a virus discovery strategy based on random amplification of purified viral nucleic acids in combination with next generation sequencing called DNase-SISPA [1]. This approach consists of a virus purification step by selective filtration and ultracentrifugation followed by nuclease treatment to remove nucleic acids that are not encapsidated in virions. This is followed by viral DNA extraction and random amplification step. The random amplicons are subsequently sequenced using an NGS technology [25].

Gudenkauf and Hewson [13] reported that distinct viral assemblages inhabit different phyla of marine invertebrates. In the present study, we hypothesized that variation in viral assemblages can also be observed in a single species of marine invertebrate from different geographic locations. This study is important in the light of the fact that wild populations still serve as the source of stocks used in aquaculture. Thus, we aim to characterize the viromes and detect putative viral pathogens by multiplexed metagenome sequencing of DNase SISPA random amplicons from the hepatopancreas of wild adult *P. monodon* from the Philippines.

Materials and methods

Sample collection and purification of virus particles

Adult wild *P. monodon* samples ($n = 4$) were purchased from local fishermen at each of 6 locations in the northern, central and southern Philippines: Camarines Norte, Quezon, Leyte, Pangasinan, Cagayan and Zamboanga del Sur. The 6 sites were traditional sources of shrimp broodstocks used for hatchery operations. Live samples were collected and frozen before transporting to the laboratory in a sealed ice box. The samples were stored in a -80°C biomedical freezer until processed.

Hepatopancreas from each sample were aseptically dissected and resuspended in 500 μl 1X phosphate buffer saline (PBS) and vigorously vortexed. Five hundred microliters of supernatant were then centrifuged at

$10,000\times g$ using a benchtop microcentrifuge for 2 min, and supernatants were transferred to a fresh tube. This step was repeated twice. The samples were stored at -80°C until further use. The pooled supernatants were then filtered through a $0.45\text{-}\mu\text{m}$ -pore-size syringe filter (Fisher Scientific, Pittsburgh, PA). The filtered supernatants were then brought up to 500 μl using 1X PBS to fill Beckman Optima™ TLX Ultracentrifuge tubes (Beckman Coulter, Inc., Brea, CA) to capacity. The pooled supernatants were then centrifuged at $50,000\times g$ for 3 h at 10°C . The pellet was resuspended in 100 μl of 1X PBS and frozen at -80°C until further processing could be performed [33].

Isolation of particle-associated nucleic acids (PAN), viral DNA extraction and whole genome amplification (WGA)

To reduce the amount of contaminating RNA and DNA present, particle-associated nucleic acids (PAN) were isolated wherein each sample (116 μl) was treated with 14 U of DNase I (Qiagen, Valencia, CA) and 20 U of RNase A (Qiagen, Valencia, CA). Samples were then brought up to a final volume of 140 μl in 10X DNase buffer (Qiagen, Valencia, CA) and incubated at 37°C for 2 h [27]. Samples were then immediately processed with GeneAll Exgene™ Viral DNA Kit (GeneAll Biotechnology Co., Ltd., Seoul, South Korea) using the manufacturer’s protocol. Since only small amounts of viral DNA were obtained due to limited samples, the samples were pooled per site and WGA was employed. GenomePlex Complete Whole Genome Amplification (WGA) Kit (Sigma-Aldrich, St Louis, MO, USA) was used following the manufacturer’s recommendations. WGA-amplified genomic material was purified using QIAquick PCR purification kit (Qiagen, Valencia, USA) and quantified using BioSpec-Nano (Shimadzu®, Kyoto, Japan).

Sequence-independent single primer amplification (SISPA) and 454 pyrosequencing

A set of highly degenerate PCR primers with barcodes as described in Table 1 was used to PCR-amplify the resulting DNA. After visualization of the random amplified DNA fragments on a 1% agarose gel, the amplicons were excised, purified and quantified using a BioSpec-Nano (Shimadzu®, Kyoto, Japan). Five-micrograms of the amplicon library was further purified with Agencourt AMPure XP beads (Beckman Coulter Inc, Canada) and quantified by fluorometry using the QuantiT™ PicoGreen® dsDNA Assay Kit (Invitrogen, Burlington, ON). The DNA fragmentation step by nebulization was skipped. Pooled amplicons were diluted as recommended, and amplified by emulsion PCR.

Table 1 Primers used in the study, adapted from Rosseel et al. [25]

Primer name	Sequence (5'–3')
FR20RV-1	<u>TTGGTTGGG</u> CCGAGCTCTGCAGATATC
FR20RV-2	<u>TTGGTTCCG</u> CCGAGCTCTGCAGATATC
FR20RV-3	<u>TTGGTAGCG</u> CCGAGCTCTGCAGATATC
FR20RV-4	<u>TTGGTACGG</u> CCGAGCTCTGCAGATATC
FR20RV-5	<u>TTGGCGTAG</u> CCGAGCTCTGCAGATATC
FR20RV-6	<u>TTGGCGATG</u> CCGAGCTCTGCAGATATC

Underlined nucleotides are barcodes used in multiplexed sequencing

Pyrosequencing was performed for 200 cycles on a Roche 454 GS-Junior sequencing instrument. All reads were filtered using the standard read rejecting filters of the built-in GS Junior sequencer software, namely key pass filters, dot and mixed filters, signal intensity filters, and primer filters. All steps were performed according to manufacturer's protocols.

Virome analysis

A Python script developed in the laboratory was used to (1) sort the sequences based on primer barcodes, (2) trim barcode and primer sequences, (3) remove homopolymers greater than 8 bp, and sequences shorter than 40 bp, allowing for 1 mismatch to the barcode and 2 mismatches to the primer, and (4) cluster sequences with a maximum e-value of 30 and a minimum percent identity of 97 as described [18].

Reads were assembled to contiguous sequences (contigs) using default parameters of Newbler *de novo* Assembler 2.7. Sequences in the NCBI Protein (<ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/nr>) and Viral Reference Sequence (RefSeq) database (<ftp://ftp.ncbi.nih.gov/refseq/release/viral>) similar to the assembled contigs were then searched using BLAST-X with an E-value cut-off of 10^{-5} [2]. The BLAST-X output was parsed using the MEta-Genome Analyzer (MEGAN) version 5.6.6 [17] (Min Score = 50.0, Max Expected = $1.0E-5$, Top Percent = 10.0, Min Support Percent = 0.1, Min Support = 1, and LCA Percent = 100.0). MEGAN was used for (1) generation of rarefaction curve, (2) determination of number of OTUs, (3) taxonomic analysis and (4) computation of Shannon-Weaver and Inverse Simpson index [18].

To comprehensively identify pathogenic virus-like sequences in our dataset, a list of viral pathogens listed by the World Organization for Animal Health (OIE) was retrieved [34]. Genetic and taxonomic information of viral pathogens of fish and shrimp were obtained from ViralZone database (<http://viralzone.expasy.org>) [16].

Phylogenetic analysis

Representative avian virus, insect virus and bacteriophage observed in *P. monodon* viromes were chosen for phylogenetic analysis. The taxonomic identity of the selected viral sequence was inferred by comparison with the closest BLAST-X match against the RefSeq Database at NCBI. Metagenomic sequence and their closest matches in GenBank were aligned using CLUSTALW and bootstrapped phylogenetic trees (1000 iterations) were generated using the maximum likelihood algorithm in MEGA6 [28].

Results and discussion

Overview of virome composition of wild *P. monodon*

Viral metagenomics was performed to characterize the virome profile and identify putative pathogenic viruses in wild *P. monodon* in the Philippines. A total of 12,405 sequence reads were obtained from all samples after quality filtering. This dataset is comparable in size to those reported by other groups using the DNase-SISPA method [11, 19]. The number of sequence reads obtained from five sites varied from 271 to 978; a dramatically larger number of reads was obtained from one site (Table 2). These differences may reflect differences in viral load although the effects of other factors (such as amount or number of cells per unit weight of tissue) cannot be fully ruled out. All sequences obtained in this study are available in the NCBI Sequence Read Archive (www.ncbi.nlm.nih.gov/sra) under accession number SRR5239623.

The sequences were further clustered into 3549 OTUs using a 97% similarity cut off [29]. Of these, between 15 and 37% matched viral genomes while the remaining sequences correspond to mostly bacterial sequences. This finding is similar to previous reports of viral metagenomics of marine invertebrates [13, 14]. The rarefaction curve generated from the OTUs is presented in Fig. 1a. Most sites were under-sampled indicating unseen diversity of viruses. However, other sites are relatively more effectively sampled as indicated by the rarefaction curve reaching plateau. Based on Shannon-Weaver and Inverse Simpson indices, the pooled sample from a site in southern Philippines (Zamboanga del Sur) had the highest virome richness while a site in northeastern Philippines (Camarines Norte) has the lowest (Table 3).

Pathogenic shrimp virus-like sequences

WSSV

WSSV, the causative agent of white spot disease (WSD) was found in samples from three of the six sites (Camarines

Table 2 Summary of results using Roche 454 GS Junior Pyrosequencing of virome, OTUs and diversity indices calculated from the sequencing reads

Collection site	No. of reads	No. of OTUs	Shannon–Weaver	Inverse simpson
Camarines Norte	9098	314	0.994	1.400
Quezon	978	945	1.000	2.000
Leyte	793	778	1.946	2.350
Pangasinan	281	271	1.663	1.943
Cagayan	419	419	3.006	7.258
Zamboanga del Sur	836	822	3.379	9.800

Fig. 1 **a** Rarefaction curve analysis of putative viral taxa in the assembled sequencing reads generated using sequence-independent single primer amplified particle-associated nucleic acids from wild *P. monodon* collected from 6 sites in the Philippines. **b** Relative diversity

Norte, Leyte and Zamboanga del Sur; Table 3, Fig. 2). WSSV is a very large, enveloped, double-stranded DNA (dsDNA) virus and considered as the most significant pathogen in the penaeid shrimp aquaculture in the world [29]. It was assigned by the International Committee on Taxonomy of Viruses (ICTV) to its own new genus, *Whispovirus*, and family, *Nimaviridae* [10].

Penaeus stylirostris densovirus (*PstDV*)

PstDV was also found in all six sites except one (Pangasinan) (Table 3; Fig. 2). *PstDV* is the smallest known penaeid virus which is 22 nm diameter, non-enveloped icosahedron with a linear single-stranded 3.9 kbp DNA. It is designated as a member of family *Parvoviridae* and a member of genus *Brevidensovirus* [10].

Other viral sequences in the *P. monodon* metagenome

Based on the most significant BLAST-X similarities, the *P. monodon* viromes contained sequences related to a wide range of avian, insect, plankton and bacterial viruses (Tables 3, 4, Figs. 1b, 2), all of which have not been previously reported to be associated with *P. monodon*, although a few have been known to infect other shrimp

of various types of viruses (classified according to their putative hosts) in the virome of wild *P. monodon* from 6 sites in the Philippines

species. The similarity of these sequences to known viral sequences (in GenBank) ranged from 27 to 100% amino acid identity. This finding is consistent with those of other viral metagenomics studies [21, 22], and suggests that wild *P. monodon* populations may harbor diverse and hitherto undiscovered, hence uncharacterized and unclassified, viral taxa. In light of this observation, continued research on virus discovery and viral genome sequencing have the potential to uncover the true diversity of viruses in wild *P. monodon*.

Variability of virome profiles in *P. monodon*

Analysis using BLAST-X of metagenomic samples from all six sites revealed the virome profile variability across the sites (Fig. 2). Samples from all sites except for one (Pangasinan) yielded sequences for the two major shrimp viruses WSSV and *PstDV*, but no sequences were found for the other important (DNA) shrimp virus such as hepatopancreatic parvovirus (HPV). (Since the procedure targets DNA, RNA viruses were not expected to be detected by the methodology). Although WSSV and *PstDV* sequences were not observed in the Pangasinan samples, a pathogenic avian virus sequence was found. Insect viral sequences were also found in samples from sites in central (Leyte), northern (Pangasinan, Cagayan) and southern

Table 3 Number of unique contigs with significant similarity (e-value < 10^{-5}) to viral pathogen sequences, obtained using BLAST-X search against NCBI Viral RefSeq database

Virus name (highest scoring hit)	Type of host	Taxonomic classification	Percent amino acid identity (%)	Length (aa)	e-value	Camarines Norte	Quezon	Leyte	Pangasinan	Cagayan	Zamboanga del Sur
White spot syndrome virus	Crustaceans	<i>Nimaviridae</i> , <i>Whispovirus</i>	52–100	26–157	1.187297×10^{-66} – 7.972527×10^{-7}	3	0	1	0	0	1
Monodon baculovirus	Penaeids	<i>Baculoviridae</i> , <i>Nudivirus</i>	27–28	31–134	3.954512×10^{-6} – 8.809733×10^{-6}	1	1	1	0	1	1
Infectious hypodermal and hematopoietic necrosis virus	Penaeids	<i>Parvoviridae</i> , <i>Brevdensovirus</i>	96–100	51–107	3.628056×10^{-52} – 1.996391×10^{-21}	1	1	1	0	1	1
Psittacine beak and feather disease virus	Old and New World parrots	<i>Circoviridae</i> , <i>Circovirus</i>	44–59	59–93	1.947310×10^{-13} – 7.399771×10^{-13}	0	0	0	1	0	0

Philippines (Zamboanga del Sur). Plankton-associated viral sequences were found in samples from Leyte and Zamboanga del Sur. Bacteriophage sequences were found in all sites except for a site in northeastern Philippines (Quezon) but almost 60% of its OTUs matched sequences for viruses with unspecified hosts (Fig. 1b).

Furthermore, only viruses that target shrimps as hosts (WSSV and *PstDV*) were commonly found in the dataset from 5 sampling sites, and sequences representing a large number of viruses with unknown hosts and those similar to other viral genomes tended to be site-specific (Table 4). This finding suggests that wild *P. monodon* metagenomes share a common but small set of core virome, particularly shrimp viruses, while the rest of the virome may be accounted for by a highly variable component consisting of other viruses such as avian, insect, plankton and bacterial viruses (Fig. 1b). Similar findings were also previously reported in mosquito viromes with shared core set of insect viruses while the remaining components tend to be more variable [21, 22].

Monodon baculovirus (MBV)-like sequences identified in *P. monodon*

Previous studies have established the presence of MBV in populations of *P. monodon* in the Philippines [20, 32]. Perhaps owing to the limited sample size, no MBV sequences were found in the metagenomes in the samples used in the study. However, MBV-like sequences were found, exhibiting 27–28% similarity to known MBV sequences. These sequences were found in all six sites except one (Pangasinan) (Table 3; Fig. 2).

Avian virus-like sequence identified in *P. monodon*

Psittacine beak and feather disease virus (PBFDV)-like sequences were found only in Pangasinan samples (Table 3; Figs. 2, 3a). PBFDV belongs to the genus *Circovirus* of the family *Circoviridae*. PBFDV is one of the smallest known animal viruses, with a simple and compact 2 kb ambisense single-stranded circular DNA genome that encodes two major genes. PBFDV was first described in Australian psittacine species in the 1970s [5]. This study is the first report of the presence of this sequence in an arthropod, in particular, shrimp, metagenome.

Insect virus-like sequences identified in *P. monodon*

Sequences that are similar to insect-associated viruses were identified in the samples from 4 of the 6 sites. These viruses are known to infect dragonflies (*Insecta: Odonata*) (Table 4; Figs. 1b, 2) and are unclassified ssDNA viruses. Three distinct but related sequences that were

Fig. 2 Virome profile of wild *P. monodon* samples from 6 sites in the Philippines

similar to dragonfly orbiculatus virus (DOV) sequences were found but only in the Pangasinan samples. DOV was reported to be an endogenous viral element (EVE) in the genome of the isopod crustacean *Armadillidium vulgare* [30]. Although DOV is known to be capable of integrating into the host DNA, the sequences observed in the study were likely derived from isolated viral particles rather than EVEs given the methodology used in this study.

Phylogenetic analysis of a representative insect virus-like sequence, dragonfly larvae associated circular virus 3 (Fig. 3b), revealed that this virus appeared to be very similar with a virus isolated from a yellow spotted dragonfly [6]. Other related sequences found in this study apparently represent viruses that were previously reported to be associated with damselflies [7, 8], feces from bats [12], badger, mongoose, otter [4] and Hawaiian red shrimp [13] but this is the first report of insect virus-like sequences associated with a penaeid shrimp.

Plankton virus-like sequences identified in *P. monodon*

Sequences bearing similarity to those of plankton viruses, namely *Diporeia* sp. - associated circular virus and

Labidocera aestiva—circovirus, were identified from the virome of *P. monodon* samples from 2 sites (Leyte and Zamboanga del Sur, respectively; Table 4, Fig. 2). *Diporeia* sp. - associated circular virus was identified from a freshwater pond in McMurdo Ice Shelf, Antarctica [35], *Procordulia grayi* and *Xanthocnemis zealandica* larvae [6] by sequencing the *Rep* protein which shares 45.62 and ~50.0% pairwise amino acid identity respectively with *Diporeia* sp.—associated circular virus. This novel circo-like virus was also isolated from *P. monodon* in Vietnam which shared 25% identity with the *Cap* protein of a *Diporeia* sp.—associated circular virus [24]. On the other hand, *Labidocera aestiva*—circovirus was identified from Tampa Bay, Florida, USA using viral metagenomics approach with BLASTp similarity (~255 aa) to a circo-like virus with an unknown host identified from free virioplankton in the Chesapeake Bay [9].

Bacteriophage-like sequences identified in *P. monodon*

Bacteriophages (viruses that infect bacteria) are the most abundant biological entities on Earth. Putative bacteriophage sequences were found in the *P. monodon* virome from all sites except for one site (Quezon). This is highly

Table 4 Number of unique contigs with significant similarity (e-value < 10⁻⁵) to viral sequences, obtained using BLAST-X search against Viral RefSeq database

Virus name (highest scoring hit)	Taxonomic classification	Percent amino acid identity (%)	Length (aa)	E-value	Camarines Norte	Quezon	Leyte	Pangasinan	Cagayan	Zamboanga del Sur
<i>Insect viruses</i>										
Dragonfly-associated circular virus 4	Unclassified ssDNA viruses	66	42	5.53332 × 10 ⁻⁸	0	0	0	1	0	0
<i>Odonata</i> -associated circular virus-9	Unclassified ssDNA viruses	36	94	2.506649 × 10 ⁻⁸	0	0	1	0	0	0
Dragonfly orbiculatus virus	No data available	30–72	36–92	6.737313 × 10 ⁻⁶ 6.517536 × 10 ⁻¹⁷	0	0	0	3	0	0
Dragonfly larvae associated circular virus 3	Unclassified ssDNA viruses	60	46	1.888731 × 10 ⁻¹³ 2.015387 × 10 ⁻¹²	0	0	0	0	1	1
<i>Odonata</i> -associated circular virus 10	Unclassified ssDNA viruses	46–45	108–81	5.020742 × 10 ⁻¹⁷ 7.217395 × 10 ⁻⁸	0	0	0	0	0	2
<i>Plankton viruses</i>										
<i>Diporeia</i> sp. – associated circular virus	Unclassified viruses	43	69	1.449990 × 10 ⁻⁸	0	0	1	0	0	0
<i>Labidocera aestiva</i> circovirus	<i>Circoviridae</i> , unclassified <i>Circovirus</i>	35	84	1.464443 × 10 ⁻⁸	0	0	0	0	0	1
<i>Bacteriophages</i>										
<i>Brochothrix</i> phage A9	<i>Myoviridae</i> ; unclassified <i>Spovnavirinae</i>	92–97	48–69	7.774100 × 10 ⁻¹⁸ 9.868536 × 10 ⁻³⁵	3	0	0	0	0	0
<i>Shewanella</i> sp. phage 3/49	<i>Myoviridae</i> , unclassified <i>Myoviridae</i>	40–56	71–50	5.162678 × 10 ⁻⁶ 3.539256 × 10 ⁻⁷	0	0	0	0	2	0
<i>Vibrio</i> phage JA-1	<i>Podoviridae</i> , N4-like virus	81	74	2.694496 × 10 ⁻²⁶	0	0	0	0	1	0
<i>Vibrio</i> phage Vc1	<i>Podoviridae</i> , unclassified <i>Podoviridae</i>	58–90	50–95	4.464848 × 10 ⁻¹⁸ 1.452595 × 10 ⁻²⁴	0	0	0	0	3	0
<i>Vibrio</i> phage pYD38-B	<i>Siphoviridae</i> , unclassified <i>Siphoviridae</i>	93	59	5.484499 × 10 ⁻²⁶	0	0	0	0	1	0
<i>Synechococcus</i> phage S-CAM8	<i>Myoviridae</i> , unclassified <i>Myoviridae</i>	81–83	31–32	1.395666 × 10 ⁻⁶ 1.046662 × 10 ⁻⁶	0	0	0	2	0	0
<i>Puniceispirillum</i> phage HMO-2011	<i>Podoviridae</i> , unclassified <i>Podoviridae</i>	84	76	1.079125 × 10 ⁻²⁷	0	0	0	1	0	0
Sp6-like virus	<i>Podoviridae</i> , SP6likevirus	82–85	34	2.356261 × 10 ⁻⁶ 3.956274 × 10 ⁻⁶	0	0	2	0	0	0
Marine gokushovirus	<i>Microviridae</i> , unclassified <i>Gokushovirinae</i>	74	81	5.363045 × 10 ⁻¹¹	0	0	0	0	0	1
<i>Unknown Host(s)</i>										
<i>Circoviridae</i> 21 LDMD 2013	<i>Circoviridae</i> , unclassified <i>Circovirus</i>	46	58	3.221802 × 10 ⁻⁸	0	0	0	0	0	1
<i>Circoviridae</i> 2013a	<i>Circoviridae</i> , unclassified <i>Circovirus</i>	39	76	3.015703 × 10 ⁻⁶	0	0	0	0	0	1

Fig. 3 Maximum likelihood tree showing the phylogenetic relationship of the **a** putative replication-associated protein of beak and feather disease virus (BFDV) sequence inferred from the virome of Pangasinan samples, **b** putative replication-associated protein of dragonfly larvae-associated circular virus-3 sequence inferred from the virome of Cagayan samples and **c** hypothetical protein of *Vibrio*

phage JA-1 sequence inferred from the virome of Cagayan samples. The tree includes highly similar sequences from GenBank based on a BLAST-X search against the non-redundant database at the NCBI. Bootstrap values of 1000 tree iterations that are > 50% are indicated in the branch nodes. BFDV Beak and feather disease virus

expected as bacteria exist within epithelial and hepatopancreatic tissues of decapods [15]. In fact, bacteriophages infecting *Brochothrix* (Camarines Norte),

Shewanella, *Vibrio* (Cagayan; Fig. 3c), *Synechococcus*, *Puniceispirillum* (Pangasinan) and Sp6-like virus (Leyte) were identified in this study (Table 4; Fig. 2).

DNase SISPA as an effective method for exploring viral diversity

By using DNase-SISPA, this study generated a baseline information of the diversity of the (DNA) virus community present in wild *P. monodon* in the Philippines. The data reflected significant viral diversity as indicated by the presence of crustacean, avian, insect, plankton and bacterial virus-like sequences, and by the fact that there are still numerous unclassified sequences which cannot be definitively identified due to limitations in the current set of viral sequences in the Viral RefSeq. Modification of the methodology to include RNA viruses would likely reveal a significantly greater diversity of the *P. monodon* virome.

The common occurrence of sequences representing shrimp pathogenic viruses confirmed their wide distributions in the wild populations of *P. monodon*, although the shrimps were apparently asymptomatic. The discovery of a high diversity of viruses in the *P. monodon* virome highlights the need to study the interactions between the pathogenic and non-pathogenic viruses.

Our study highlighted the advantages of the DNase-SISPA method in characterizing the diversity of the *P. monodon* virome and identifying pathogenic and other associated viruses which are likely to be missed by conventional methods. In particular, DNase-SISPA can give a snapshot of the diversity of the viruses present in a sample unlike conventional methods which are generally designed to target specific viruses. Shrimp samples from different regions displayed variability of virome profiles. Since the identity of other potential viruses have been revealed by this study, these sequences can now serve as marker that will be useful in investigating their potential involvement in occurrences or outbreaks of diseases. Moreover, improvement of the methodology to target RNA viruses with higher sample size and using deep sequencing technologies will be necessary to gain a more comprehensive view of the *P. monodon* virome diversity.

Acknowledgements Special thanks to DT Boncan, BT Torres, IC Uy, ZT Aquinaldo, JA Nieva, ME Dela Cruz, LF Cunanan, EC Eugenio and AE David of Marine Genomics and Molecular Genetics Laboratory for the help in collection of samples and laboratory experiments, to Ms. Abigail Albaladejo of Bureau of Fisheries and Aquatic Resources (BFAR) for helping in sample collection, to Department of Science and Technology—Philippine Council for Agriculture, Aquatic and Natural Resources Research and Development (DOST – PCAARRD) for funding the project under AOL, and to Department of Science and Technology – Advanced Science and Technology Human Resource Development Program (DOST – ASTHRDP) for the scholarship grant to FLO.

References

1. Allander T, Emerson SU, Engle RE, Purcell RH, Bukh J. A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species. *Proc Natl Acad Sci*. 2001;98:11609–14.
2. Altschul S, Gish W, Miller W, Myers E. Basic local alignment search tool. *J Mol Biol*. 1990;215:403–10.
3. Bidle KD. Elucidating marine virus ecology through a unified heartbeat. *Proc Natl Acad Sci*. 2014;111:15606–7.
4. Conceição-Neto N, Zeller M, Heylen E, Lefrère H, Mesquita JR, Matthijssens J. Fecal virome analysis of three carnivores reveals a novel nodavirus and multiple gemycircularviruses. *Virol J*. 2015;12:79.
5. Crowther R, Berriman J, Curran W. Comparison of the structures of three circoviruses: chicken anemia virus, porcine circovirus type 2, and beak and feather disease virus. *J Virol*. 2003;77:13036–41.
6. Dayaram A, Galatowitsch M. Diverse circular replication-associated protein encoding viruses circulating in invertebrates within a lake ecosystem. *Infect Genet Evol*. 2016;39:304–16.
7. Dayaram A, Galatowitsch M, Harding JS, Argüello-Astorga GR, Varsani A. Novel circular DNA viruses identified in *Procordulia grayi* and *Xanthocnemis zealandica* larvae using metagenomic approaches. *Infect Genet Evol*. 2014;22:134–41.
8. Dayaram A, Potter KA, Pailles R, Marinov M, Rosenstein DD, Varsani A. Identification of diverse circular single-stranded DNA viruses in adult dragonflies and damselflies (*Insecta: Odonata*) of Arizona and Oklahoma, USA. *Infect Genet Evol*. 2015;30:278–87.
9. Dunlap DS, Ng TFF, Rosario K, Barbosa JG, Greco AM, Breitbart M, Hewson I. Molecular and microscopic evidence of viruses in marine copepods. *Proc Natl Acad Sci*. 2013;110:1375–80.
10. Fauquet C, Mayo M, Maniloff J. Virus taxonomy: VIIIth report of the international committee on taxonomy of viruses. Cambridge: Academic Press; 2005.
11. Finkbeiner S, Allred A, Tarr P. Metagenomic analysis of human diarrhea: viral detection and discovery. *PLoS Path*. 2008;4:e1000011.
12. Ge X, Li J, Peng C, Wu L, Yang X, Wu Y, Zhang Y, Shi Z. Genetic diversity of novel circular ssDNA viruses in bats in China. *J Gen Virol*. 2011;92:2646–53.
13. Gudenkauf B, Hewson I. Comparative metagenomics of viral assemblages inhabiting four phyla of marine invertebrates. *Front Mar Sci*. 2016;3:23.
14. Gudenkauf BM, Eaglesham JB, Aragundi WM, Hewson I. Discovery of urchin-associated densoviruses (family *Parvoviridae*) in coastal waters of the Big Island, Hawaii. *J Gen Virol*. 2014;95:652–8.
15. Guri M, Durand L, Cueff-Gauchard V, Zbinden M. Acquisition of epibiotic bacteria along the life cycle of the hydrothermal shrimp *Rimicaris exoculata*. *The ISME J*. 2012;6:597–609.
16. Hulo C, De Castro E. ViralZone: a knowledge resource to understand virus diversity. *Nucleic Acids Res*. 2011;39:D576–82.
17. Huson D, Mitra S, Ruscheweyh H. Integrative analysis of environmental sequences using MEGAN4. *Genome Res*. 2011;21:1552–60.
18. Kim Y, Aw T, Teal T, Rose J. Metagenomic investigation of viral communities in ballast water. *Environ Sci Technol*. 2015;49:8396–407.
19. Lysholm F, Wetterbom A, Lindau C. Characterization of the viral microbiome in patients with severe lower respiratory tract infections, using metagenomic sequencing. *PLoS ONE*. 2012;7:e30875.

20. Natividad K, Migo M. Simultaneous PCR detection of two shrimp viruses (WSSV and MBV) in postlarvae of *Penaeus monodon* in the Philippines. *Aquaculture*. 2006;257:142–9.
21. Ng T, Manire C, Borrowman K. Discovery of a novel single-stranded DNA virus from a sea turtle fibropapilloma by using viral metagenomics. *J Virol*. 2009;83:2500–9.
22. Ng T, Willner D, Lim Y, Schmieder R. Broad surveys of DNA viral diversity obtained through viral metagenomics of mosquitoes. *PLoS ONE*. 2011;6:e20579.
23. Park E, Kim K, Abell G. Metagenomic analysis of the viral communities in fermented foods. *Appl Environ Microbiol*. 2011;77:1284–91.
24. Pham HT, Yu Q, Boisvert M, Van HT, Bergoin M, Tijssen P. A circo-like virus isolated from *Penaeus monodon* shrimps. *Genome Announc*. 2014;2:e01172-13.
25. Rosseel T, Scheuch M, Höper D. DNase SIPA-next generation sequencing confirms Schmallenberg virus in Belgian field samples and identifies genetic variation in Europe. *PLoS ONE*. 2012;7:e41967.
26. Schoenfeld T. Assembly of viral metagenomes from yellowstone hot springs. *Appl Environ Microbiol*. 2008;74:4164–74.
27. Stang A, Korn K, Wildner O. Characterization of virus isolates by particle-associated nucleic acid PCR. *J Clin Microbiol*. 2005;43:716–20.
28. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol and Evol*. 2013;30:2725–9.
29. Tan Y, Xing Y, Zhang H, Feng Y, Zhou Y, Shi ZL. Molecular detection of three shrimp viruses and genetic variation of white spot syndrome virus in Hainan province, China, in 2007. *J Fish Dis*. 2009;32:777–84.
30. Thézé J, Leclercq S, Moumen B, Cordaux R, Gilbert C. Remarkable diversity of endogenous viruses in a crustacean genome. *Genome Biol Evol*. 2014;6:2129–40.
31. Tyson G, Chapman J, Hugenholtz P. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature*. 2004;428:37–43.
32. ickers J, Webb R, Young P. Monodon baculovirus from Australia: ultrastructural observations. *Dis Aquat Organ*. 2000;39:169–76.
33. Victoria J, Kapoor A, Li L, Blinkova O. Metagenomic analyses of viruses in stool samples from children with acute flaccid paralysis. *J Virol*. 2009;83:4642–51.
34. Walker PJ, Winton JR. Emerging viral diseases of fish and shrimp. *Vet Res*. 2010;41:51.
35. Zawar-Reza P, Argüello-Astorga GR, Kraberger S, Julian L, Stainton D, Broady PA, Varsani A. Diverse small circular single-stranded DNA viruses identified in a freshwater pond on the McMurdo Ice Shelf (Antarctica). *Infect Genet Evol*. 2014;26:132–8.