ABSTRACT: Fisheries and aquaculture are impacted sporadically by newly emerged viral diseases. In most cases, searches for a causative pathogen only occur after a serious disease has emerged. As random shotgun sequencing (metagenomics) offers opportunities to identify novel viruses preemptively, the method was tested on nucleic acids extracted from the hepatopancreas of 12 healthy northern pink shrimp *Farfantepenaeus duorarum* captured from the Gulf of Mexico. Among the sequences, a nodavirus (*Farfantepenaeus duorarum* nodavirus, *FdNV*) and a virus with similarities to circoviruses and cycloviruses that possess circular single-stranded DNA (ssDNA) genomes, were identified. The *FdNV* genome sequence was most closely related phylogenetically to nodaviruses causing white tail disease in *Macrobrachium rosenbergii* and muscle necrosis disease in *Litopenaeus vannamei*. While the circular ssDNA virus represents the third to be detected in association with a marine invertebrate, transmission trials are needed to confirm its infectivity for *F. duorarum*. This study highlights the potential for using metagenomic approaches in fisheries and aquaculture industries to identify new potential pathogens in asymptomatic marine invertebrates, uncharacterized pathogens causing a new disease, or multiple pathogens associated with disease syndromes.

KEY WORDS: Fisheries · Pathogen discovery · Metagenomics · Nodavirus · Circovirus

INTRODUCTION

In 2010, global aquaculture of over 600 species including finfish, crustaceans, and mollusks produced approximately 60 million tons of seafood with an estimated value of US $119 billion (FAO 2011). However, aquaculture worldwide has become increasingly susceptible to new diseases due mainly to movement of live animals, the increased use of high-density farming systems, and increasing anthropogenic pressures on aquatic ecosystems (FAO 2011). With tropical shrimp aquaculture, it has been esti-
mated that up to 40% of annual production (>3 billion) is currently being lost, mainly due to viral diseases (Stentiford et al. 2012).

Although emerging diseases often cause substantial economic losses, little effort is generally expended on preemptively identifying potential pathogens before they cause a problem to aquaculture. This delays development of molecular diagnostic tests for identifying pathogen reservoirs and vectors, host and geographic ranges, modes of transmission, and genetic relatedness to other pathogens, impeding our understanding of disease epidemiology and ability to minimize disease impacts (Walker & Winton 2010). The reactive approach also neglects the fact that many potential pathogens have evolved to coexist in symbiosis with particular hosts, with transition to disease states only occurring after some ecological imbalance promoted by environmental or anthropogenic factors (Walker & Winton 2010). Understanding which potential pathogens exist naturally in commercially important species would thus assist preparedness for preventing or managing disease should a pathogen emerge as a problem.

Various methods are available for identifying and characterizing novel microorganisms. Metagenomic sequencing of nucleic acids extracted from fluids or tissue homogenates filtered or enriched to concentrate and semi-purify virus particles has proved useful in identifying genome sequences of novel viruses in humans and various domestic and wild animals (Breitbart et al. 2002, Allander et al. 2005, Delwart 2007, Victoria et al. 2008, Ng et al. 2009, 2011a,b, Svarka et al. 2010). Once sequence data are available, options abound to develop sensitive, specific, and rapid molecular tests for virus diagnosis and epidemiology.

Due to the wild shrimp fishery in southern parts of the Gulf of Mexico showing signs of being depleted, aquaculture of northern pink shrimp Farfantepenaeus duorarum has become increasingly important to the economy of the region (Arreguin-Sanchez et al. 2008). To avoid disease impacts, pathogen-free breeding stocks of F. duorarum have also been sought to support the industry (Samocha et al. 2008). To investigate whether viruses might exist in healthy F. duorarum indigenous to the Gulf of Mexico that could pose a potential risk to aquaculture, random shotgun sequencing was undertaken on extracts of hepatopancreas tissue. This led to the identification of a new nodavirus and a virus with similarities to circoviruses and cycloviruses that possess circular single-stranded DNA (ssDNA) genomes.

**MATERIALS AND METHODS**

Overly healthy juvenile Farfantepenaeus duorarum (6–8 g) caught near Tarpon Springs, Florida, USA, in the Gulf of Mexico, were transported to the laboratory in aerated buckets. Hepatopancreas tissue dissected aseptically from 12 shrimp was pooled and homogenized in sterile SM buffer (50 mM Tris, 10 mM MgSO4, 0.1 M NaCl, pH 7.5) using a Tissumizer (Tekmar). Virus particles were purified from filtered homogenate as described previously (Breitbart & Rohwer 2005, Ng et al. 2011a, 2012). Briefly, homogenate was clarified by centrifugation at 10,000 × g (10 min), filtered through a 0.22 µm filter, mixed with 0.2 volume chloroform for 10 min, and then incubated with 2.5 U DNase I and 0.25 U RNase A per µl at 37°C for 3 h. DNA and RNA were extracted from purified material using the Qiagen DNeasy Blood & Tissue Kit and RNeasy Mini Kit, respectively.

RNA and DNA were sequenced using cDNA synthesis, DNA amplification, cloning, and sequencing methods described previously (Breitbart & Rohwer 2005, Ng et al. 2011a). Briefly, DNA was amplified using Phi29 DNA polymerase (GenomiPhi, GE Healthcare), fragmented and amplified again using a Whole Genome Amplification kit (Sigma-Aldrich). RNA was converted to cDNA and amplified by PCR using a TransPlex Whole Transcriptome Amplification kit (Sigma-Aldrich). Randomly amplified DNA libraries were cloned into a TOPO TA vector (Invitrogen) and transformed into competent cells. Inserts in 50 to 130 clones from the DNA and RNA libraries were amplified by PCR using M13 primers, and amplicons >150 bp were Sanger sequenced, trimmed, and assembled into contigs applying a match size = 35, minimum match percentage = 95% using the SeqMan Pro-assembler (DNASTAR). Singlets and assembled contigs were compared to sequences in the GenBank non-redundant database using BLASTX (Altschul et al. 1990, 1997), leading to the identification of a circovirus-like DNA sequence and a nodavirus-like RNA sequence.

To generate a complete circovirus-like genome sequence by inverse PCR (Ng et al. 2009, 2011a), DNA was amplified randomly using Phi29 DNA polymerase (TempliPhi, GE Healthcare) before PCR for 45 cycles employing the outward facing primers 5′TGA CAT TGG GAT ACC ACT GG 143-3′ and 5′TCA AGG ATA CTG CTG CCA TG 107-3′, the ~2 kb DNA amplified by PCR was cloned and Sanger sequenced by primer walking.

Amino acid sequences of the circovirus-like replication initiator protein (Rep) and nodavirus capsid protein (Cap) were aligned to various homologues
using MUSCLE (Edgar 2004) with manual editing. The multiple alignments were used to infer maximum-likelihood phylogenetic trees using PHYML Version 3.0 (Guindon et al. 2010) with approximate likelihood-ratio test (aLRT) branch support (Anisimova & Gascuel 2006) and an LG model of substitution. Mesquite (version 2.75) was used to collapse branches with <60% aLRT branch support.

RESULTS AND DISCUSSION

Virus particles were partially purified from a homogenate of hepatopancreas tissue from 12 wild *Farfantepenaeus duorarum* using filtration, chloroform, and nuclease treatment. Sequence analysis of clones derived from either RNA or DNA extracted and amplified from this material identified a new DNA virus possessing a circular genome designated shrimp hepatopancreas-associated circular DNA virus (ShrimpCDV) and a new nodavirus designated *Farfantepenaeus duorarum* nodavirus (FdNV).

**ShrimpCDV**

The 1956 nt ssDNA genome of ShrimpCDV (Fig. 1A; GenBank Accession KC441518) contains 2 genes encoding putative Rep and Cap proteins transcribed bi-directionally. A putative DNA hairpin structure containing a sequence (AGG TAT TAC) similar to the conserved nonanucleotide motif of circoviruses exists in the short intergenic region (SIR; Fig. 1A). In pairwise distance analyses, the ShrimpCDV Rep protein was 21 to 34% identical to the cognate protein of circoviruses, cycloviruses, and other unclassified ssDNA viruses. Fig. 1B illustrates the phylogenetic relationships between ShrimpCDV and other circular ssDNA viruses. According to the classification scheme based on genome organization developed recently for circular ssDNA viruses (Rosario et al. 2012), ShrimpCDV is a Type IV virus.

Numerous novel circular ssDNA virus genomes have been identified recently by viral metagenomics from insects, fecal matter, and various environmental samples, including, for example, reclaimed water and sewage (Delwart & Li 2012, Rosario et al. 2012). Although many genomes have been identified in seawater and one was found associated with fish (Rosario et al. 2009, Lorincz et al. 2011), with regard to marine invertebrates, circular ssDNA viruses have only recently been found associated with 2 copepod species (*Acartia tonsa* and *Labidocera aestiva*; Dunlap et al. 2013). Due to their prevalence in the environment, it is not unexpected that circular ssDNA viruses should be found associated with marine organisms. Since the hepatopancreas is a digestive organ of shrimp, and semi-purified viral DNA extracted from the hepatopancreas tissue was amplified using Phi29 DNA polymerase followed by random PCR, it cannot be discounted that ShrimpCDV represents an ingested environmental contaminant resilient to degradation by the shrimp digestive system. To gain more knowledge regarding the host of ShrimpCDV, molecular screening of wild *F. duorarum* needs to be undertaken to verify the presence of this virus in shrimp tissues exclusive of the digestive track, and bioassays should be performed to test its ability to infect *F. duorarum*.

**FdNV**

Two nodavirus-like sequences were identified among clones generated from the *Farfantepenaeus duorarum* hepatopancreas RNA virome. Nodaviruses are small, spherical, non-enveloped viruses containing a genome comprised of 2 positive sense ssRNAs (Scherer & Hurlbut 1967, Scotti et al. 1983, Dasmahapatra et al. 1985, Reinganum et al. 1985, Zeddam et al. 1999), with the larger RNA1 encoding a RNA-dependent RNA polymerase and the smaller RNA2 encoding the structural capsid protein (Friesen & Rueckert 1981, Ball 1995, Nagai & Nishizawa 1999). The *Nodaviridae* includes insect-infecting viruses classified in the genus *Alphanodavirus* (Scherer & Hurlbut 1967, Scotti et al. 1983, Dasmahapatra et al. 1985, Reinganum et al. 1985, Zeddam et al. 1999), fish-infecting viruses classified in the genus *Betanodavirus* (Munday et al. 2002), and the as yet unclassified *Macrobrachium rosenbergii* nodavirus (MrvNV) and *Penaeus vannamei* nodavirus (PvNV; from *Litopenaeus vannamei*) that cause muscle diseases in these shrimp species (Arcier et al. 1999, Tang et al. 2007).

In our study, 2 clones were identified from the *Farfantepenaeus duorarum* hepatopancreas tissue that shared amino acid identities to previously described nodaviruses. One clone contained a 403 nt insert encoding an RNA-dependent RNA polymerase (RdRp) partial sequence (GenBank Accession: KC441519) and the other contained a 236 nt insert encoding a capsid protein partial sequence (GenBank Accession: KC441520). Phylogenetic analysis using the partial sequence of the FdNV capsid protein, which is used to classify nodaviruses, revealed 43 to 51% amino acid identity to the 2 previously characterized
Whether FdNV has potential to cause muscle disease in *F. duorarum* similar to the diseases described for MrNV and PvNV has yet to be determined. However, the genome sequence information generated through the metagenomic approach used here will allow sensitive molecular tests to be developed to assist in understanding, among other things, the potential pathogenicity of FdNV for *F. duorarum*.

The discovery of a new nodavirus in healthy *Farfantepenaeus duorarum* adds to the picture of nodavirus diversity among crustaceans and highlights the risks of translocation of unknown viruses of live crustaceans. For example, MrNV and PvNV now occur in both Western and Eastern Hemisphere regions where these species are cultured as the result of trade in live crustaceans (Qian et al. 2003, Yogananth et al. 2006). The low genetic relatedness of FdNV to MrNV and PvNV (Fig. 2) suggests that these viruses have evolved over long periods in their preferred host species, and that metagenomic approaches such as those used here might find similarly divergent nodaviruses in other shrimp species.
Metagenomics

The use of a metagenomic approach to identify a new nodavirus and a novel circular ssDNA virus in overtly healthy *Farfantepenaeus duorarum* highlights the potential value of such methods for discovering unknown viruses in commercially-important fisheries and aquaculture species. Such preemptive discoveries offer opportunities to develop sensitive molecular diagnostic tests and to guide research to determine the potential threat of newly-discovered viruses in emerging as a significant disease (Walker & Winton 2010). Metagenomic approaches can similarly be used to identify undiagnosed pathogens causative of newly-emerged diseases (Allander et al. 2005, Delwart 2007, Victoria et al. 2008, Ng et al. 2009, 2011a, Svraka et al. 2010), thereby expediting development of diagnostic tools to help prevent or manage these diseases.

Acknowledgements. This work was funded through grants DEB-1239976 and OCE-1049670 from the National Science Foundation to M.B. T.N. was funded by the William and Elsie Knight Oceanographic Fellowship. S.V.A. was supported by the DBT Overseas Fellowship from the Department of Biotechnology, Government of India. We thank B. Dwivedi for assistance with bioinformatics, S. Habtes and E. Peebles for identification of shrimp, and T. MacDonald for scientific discussion.

LITERATURE CITED


Delwart E, Li L (2012) Rapidly expanding genetic diversity and host range of the Circoviridae family and other Rep encoding small circular ssDNA genomes. Virus Res 164:114–121


FAO (Food and Agriculture Organization of the United Nations) (2011) The state of the world fisheries and aquaculture 2010. FAO, Rome


Ng TFF, Willner DL, Lim YW, Schmieder R and others (2011b) Broad surveys of DNA viral diversity obtained through viral metagenomics of mosquitoes. PLoS ONE 6:e20579


Editorial responsibility: Craig Carlson, Santa Barbara, California, USA

Submitted: October 12, 2012; Accepted: June 11, 2013
Proofs received from author(s): July 31, 2013