Infection, Genetics and Evolution 31 (2015) 284–295

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Diverse small circular DNA viruses circulating amongst estuarine molluscs

Anisha Dayaram [a], Sharyn Goldstien [a], Gerardo R. Argüello-Astorga [b], Peyman Zawar-Reza [c], Christopher Gomez [c], Jon S. Harding [a], Arvind Varsani [a,d,e,f,*]

[a] School of Biological Sciences, University of Canterbury, Christchurch 8140, New Zealand
[b] División de Biología Molecular, Instituto Potosino de Investigación Científica y Tecnológica, Camino a la Presa San José 2055, 78216 San Luis Potosí, S.L.P., Mexico
[c] Department of Geography, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand
[d] Biomolecular Interaction Centre, University of Canterbury, Christchurch 8140, New Zealand
[e] Electron Microscope Unit, Division of Medical Biochemistry, Department of Clinical Laboratory Sciences, University of Cape Town, Rondebosch 7700, South Africa
[f] Department of Plant Pathology and Emerging Pathogens Institute, University of Florida, Gainesville, FL 32611, USA

ARTICLE INFO

Article history:
Received 24 December 2014
Received in revised form 6 February 2015
Accepted 11 February 2015
Available online 17 February 2015

Keywords:
Circular DNA viruses
CRESS DNA viruses
Molluscs
Austrovenus stutchburyi
Paphies subtriangulata
Amphibola crenata

ABSTRACT

Our understanding of the diversity and abundance of circular replication associated protein (Rep) – encoding single stranded (CRESS) DNA viruses has increased considerably over the last few years due to a combination of modern sequencing technologies and new molecular tools. Studies have used these to identify and recover CRESS DNA viruses from a range of different marine organisms, including copepods, shrimp and molluscs. In our study we identified 79 novel CRESS DNA viruses from three mollusc species (*Austrovenus stutchburyi*, *Paphies subtriangulata* and *Amphibola crenata*) and benthic sediments from the Avon-Heathcote estuary in Christchurch, New Zealand. The genomes recovered have varying genome architectures, with all encoding at least two major ORFs that have either unidirectional or bidirectional organisation. Analysis of the Reps of the viral genomes showed they are all highly diverse, with only one Rep sequence sharing 65% amino acid identity with the Rep of gastropod-associated circular DNA virus (GaCSV). Our study adds significantly to the wealth of CRESS DNA viruses recovered from freshwater and marine environments and extends our knowledge of the distribution of these viruses.

1. Introduction

A large number of novel circular replication associated protein (Rep) encoding single stranded DNA (CRESS) viruses have been discovered in the last five years in various ecosystems. This is primarily due to new sequence-independent amplification methods and through the use of next generation sequencing (NGS) approaches. Many studies have used a combination of rolling circle amplification (RCA), shotgun sequencing and NGS to identify CRESS DNA viruses from a range of environments including deep-sea vents (Yoshida et al., 2013), Antarctic lakes and ponds (López-Bueno et al., 2009; Zawar-Reza et al., 2014), reclaimed water (Rosario et al., 2009b), rice paddy soil (Kim et al., 2008) and oceans (Angly et al., 2006; Labonte and Suttle, 2013; Rosario and Breitbart, 2011; Rosario et al., 2009a). Given that viruses are the most abundant entities in most environments and play an important role in regulating the structure of microbial communities (Danovaro et al., 2011; Suttle, 2007), it is not surprising that diverse CRESS DNA viruses are abundant. This raises question of their role and 'flow' within ecosystems.

Herpes-like viruses were first identified in adult *Crassostrea virginica* (Farley, 1976; Farley et al., 1972) and subsequently herpes-like viruses have been associated with high mortalities of *Crassostrea gigas* hatchery-reared larvae (Hine et al., 1992; Nicolas et al., 1992). Experimental studies have demonstrated the transmission of these herpes-like viruses between different species of molluscs and their larvae (Le Deuff et al., 1994; LeDeuff et al., 1996). Viruses belonging to the families *Birnaviridae*, *Picornaviridae*, *Togaviridae*, *Reoviridae* and *Retroviridae* have also been found to infect molluscs, reviewed by Renault and Novoa (2004).

In general, molluscs have inefficient metabolic systems and are herbivorous filter feeders and hence they tend to concentrate entities present in their aquatic environment. For this reason they can be used as indicators of contaminants and for monitoring pollutants in ecosystems (Baršienė et al., 2002; Gunkel and Streit,

* Corresponding author at: School of Biological Sciences, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand. Tel.: +64 3 366 7001x4667; fax: +64 3 364 2590.
E-mail address: arvind.varsani@canterbury.ac.nz (A. Varsani).

http://dx.doi.org/10.1016/j.meegid.2015.02.010
1567-1348/

1980; van der Oost et al., 1988; Walsh et al., 1994). The concept of bio-concentration has also been applied to areas of viral research, particularly for commercial bivalve molluscs such as oysters, clams, mussels and scallops to monitor viruses that are pathogenic to humans (Cheng et al., 2005; Comps, 1988; Elston, 1997; Farley, 1976; Meyers et al., 2009; Morley, 2010).

Recently the first CRESS DNA virus associated with the gastropod *Amphibola crenata* was described in Christchurch, New Zealand (Dayaram et al., 2013c). Gastropod-associated circular DNA virus (GaCSV) has a 2351 nt genome with two major open reading frames (ORFs) that are bidirectionally transcribed. Phylogenetic analysis of the Rep of GaCSV shows that it clusters with Rep-like sequences of bacterial origins. At the same site a *Starling circovirus* (StCV) was also identified in *A. crenata* (Dayaram et al., 2013a). Prior to this study there had been no reports of starling circoviruses outside of Europe. Starlings have been seen foraging around the site where *A. crenata* were sampled and it is highly likely that the source of the StCV was starling faecal matter. This shows a unique example that molluscs may be important 'natural tools' for viral surveillance in ecosystems.

Given that most studies have looked at specific samples within ecosystems for assessing viral diversity, we sampled *Austrovenus stutchburyi* (tuangi or cockles), *Paphies subtriangulata* (tuatua or clam) and *Amphibola crenata* (Titiko or mud-flat snail) as potential 'natural surveillance tools for viruses. We also sampled benthic sediments that are the common medium that these four species inhabit. Our objective was to identify CRESS DNA viruses circulating in the Avon-Heathcote estuary and to determine whether the mollusc species sampled in this study could be used for future viral surveillance work.

2. Materials and methods

2.1. Sample collection and processing

Samples were collected at three different sites in the Avon-Heathcote estuary on the 4th of July 2012. The estuary forms the confluence of two rivers (the Avon and Heathcote) which flow through the city of Christchurch (approx. population 340,000 people). The two rivers have >90 small tributary streams and almost their entire watersheds drain from the urban area. Fifty-five individual *A. stutchburyi* were collected from the Heathcote Bridge (near the mouth of the Heathcote River, 43.5578S, 172.70588E), 40 individuals from the Causeway (43.5564S, 172.7289E), 40 individuals from Beachville Rd (43.5567S, 172.7350E) and 40 individuals from PP the Yacht club (43.5661S, 172.7469E) (Fig. 1). Forty *P. subtriangulata* were collected from PP Yacht club (43.5661S, 172.7469E) and 50 *A. crenata* were collected from Heathcote Bridge (43.5578S, 172.7058E). Samples were refrigerated upon collection. All mollusc samples were removed from the shells and washed in sterile distilled water prior to processing. The whole bodies were then pooled to individual species for each of the sites sampled.

2.2. DNA extraction and viral purification and circular DNA enrichment

Each pooled species and the benthic sediments were homogenised in SM buffer [0.1 M NaCl, 50 mM Tris/HCl (pH 7.4), 10 mM $MgSO_4$] at a ratio of 10 ml SM buffer to 5 g of tissue. The homogenate was then centrifuged (10,000*g* for 10 min) in order to pellet the cellular debris and the supernatant was filtered first through a 0.45 μm, followed by 0.2 μm syringe filters (Sartorius Stedim Biotech, Germany). Viral DNA was then recovered using the High Pure Viral Nucleic Acid kit (Roche, USA) from the filtrate. In order to enrich circular DNA, we used rolling circle amplification (RCA) using TempliPhi 2000 (GE Healthcare, USA) as previously described (Dayaram et al., 2014, 2013c; Rosario et al., 2012a, 2011; Sikorski et al., 2013a).

2.3. Illumina HiSeq 2000 sequencing, NGS data assembly and recovery of viral genomes

The RCA products were then pooled by species and sequenced on an Illumina HiSeq 2000 (Illumina, USA) platform at the Beijing Genomics Institute (Hong Kong). The paired end reads were *de novo* assembled using ABySS V1.3.5 (Simpson et al., 2009) with kmer = 64. A full BLASTx analysis was performed on the *de novo* assembled contigs >1500 nts using KoriBlast v 4.1 (Korilog SARL, Bioinformatics Solutions, France) in order to identify contigs encoding known viral-like proteins.

Back-to-back primers were then designed (Supplementary Table 1) based on the BLASTx hits of the *de novo* assembled contigs which had similarities to CRESS DNA viral-like sequences in order to recover the full genomes from the three mollusc species and the benthic sediment samples. The genomes were recovered by PCR amplification with the specific back-to-back primers pairs using Kapa HiFi HotStart polymerase (Kapa Biosystems, USA). The resulting amplicons were gel purified, cloned into pJET1.2 plasmid (Thermo Fisher, USA) and Sanger sequenced at Macrogen Inc. (South Korea) by primer walking.

Sanger sequencing reads were assembled using DNA Baser Sequence Assembler (version 4.16; Heracle Biosoft S.R.L., Romania). A initial comparison of the genomes sequences was then carried out using BLASTx and tBLASTx (Altschul et al., 1990). The viral genomes were then annotated and the major open reading frames (ORFs), conserved motifs and stem-loop structures were identified using the Sfold server (http://sfold.wadsworth.org/cgi-bin/index.pl).

2.4. Analysis of the major ORFs

The Reps from all CRESS DNA viruses available in GenBank (downloaded 20th August 2014) were used to create a dataset for analysis. These Rep sequences together with those of viruses determined in this study were aligned using MUSCLE (Edgar, 2004) with manual editing where necessary. The resulting alignment was used to infer an approximate-maximum likelihood phylogenetic tree using a JTT + CAT model with aLRT branch (Anisimova and Gascuel, 2006) support using FastTree version 2.1.7 (Price et al., 2010). The Rep maximum likelihood phylogenetic tree was mid-point rooted and branches with less than 80% aLRT support were collapsed using Mesquite (version 2.75; http://mesquiteproject.org/).

Rep-like sequences found in bacterial genomes were downloaded from GenBank on the 5th of October 2014 and small dataset was created with the Rep of GasCV (KC172652) and that of a virus identified in this study which is most closely related to it. These Rep sequences were aligned using MUSCLE (Edgar, 2004). ProtTest (Darriba et al., 2011) was used to determine the model of best fit and the phylogenetic tree was generated using PhyML version 3.0 (Guindon et al., 2010). The Rep maximum likelihood phylogenetic tree was inferred using RtREV + G + F amino acid substitution model with aLRT branch support and was rooted with circovirus sequences, namely *Beak and feather disease virus* (BFDV; AF071878), *Porcine circovirus* (PCV1; AF012107), *Duck circovirus* (DuCV; DQ100076) and *Raven circovirus* (RaCV; DQ146997).

The Reps and capsid proteins (CPs) of ten viruses from this study with similarities to pig-stool-associated circular ssDNA virus (PisaCV), bovine stool associated circular virus (BoSVC), porcine-stool-associated circular virus (PoSCV), turkey-stool-associated circular virus (TuSCV), chimpanzee-stool-associated circular virus