Infection, Genetics and Evolution 39 (2016) 304–316

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Research paper

Diverse circular replication-associated protein encoding viruses circulating in invertebrates within a lake ecosystem

Anisha Dayaram [a], Mark L. Galatowitsch [a,b], Gerardo R. Argüello-Astorga [c], Katherine van Bysterveldt [a], Simona Kraberger [a], Daisy Stainton [a], Jon S. Harding [a], Philippe Roumagnac [d], Darren P. Martin [e], Pierre Lefeuvre [f], Arvind Varsani [a,g,h,i,*]

[a] School of Biological Sciences, University of Canterbury, Christchurch 8140, New Zealand
[b] Department of Biology, Centre College, Danville, KY 40422, USA
[c] 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
[d] Computational Biology Group, Institute of Infectious Diseases and Molecular Medicine, University of Cape Town, Cape Town 7925, South Africa
[e] CIRAD-INRA-SupAgro, UMR BGPI, Campus International de Montferrier-Baillarguet, 34398 Montpellier Cedex-5, France
[f] CIRAD, UMR PVBMT, Pôle de Protection des Plantes, 7 chemin de l'IRAT, Saint-Pierre, Ile de la Réunion 97410, France
[g] Biomolecular Interaction Centre, University of Canterbury, Christchurch 8140, New Zealand
[h] Structural Biology Research Unit, Division of Medical Biochemistry, Department of Clinical Laboratory Sciences, University of Cape Town, Rondebosch 7700, South Africa
[i] Department of Plant Pathology and Emerging Pathogens Institute, University of Florida, Gainesville, FL 32611, USA

ARTICLE INFO

Article history:
Received 26 November 2015
Received in revised form 30 January 2016
Accepted 7 February 2016
Available online 10 February 2016

Keywords:
CRESS DNA viruses
Replication-associated protein
Freshwater invertebrates
Echyridella menzeisi
Procordulia grayi
Xanthocnemis zealandica
Musculium novazelandiae
Chironomus zealandicus
Potamopyrgus antipodarum
Physella acuta

ABSTRACT

Over the last five years next-generation sequencing has become a cost effective and efficient method for identifying known and unknown microorganisms. Access to this technique has dramatically changed the field of virology, enabling a wide range of environmental viral metagenome studies to be undertaken of organisms and environmental samples from polar to tropical regions. These studies have led to the discovery of hundreds of highly divergent single stranded DNA (ssDNA) virus-like sequences encoding replication-associated proteins. Yet, few studies have explored how viruses might be shared in an ecosystem through feeding relationships. Here we identify 169 circular molecules (160 CRESS DNA molecules, nine circular molecules) recovered from a New Zealand freshwater lake, that we have tentatively classified into 51 putatively novel species and five previously described species (DflaCV-3, -5, -6, -8, -10). The CRESS DNA viruses identified in this study were recovered from molluscs (*Echyridella menzeisii*, *Musculium novaezelandiae*, *Potamopyrgus antipodarum* and *Physella acuta*) and insect larvae (*Procordulia grayi*, *Xanthocnemis zealandica*, and *Chironomus zealandicus*) collected from Lake Sarah, as well as from the lake water and benthic sediments. Extensive diversity was observed across most CRESS DNA molecules recovered. The putative capsid protein of one viral species was found to be most similar to those of members of the Tombusviridae family, thus expanding the number of known RNA–DNA hybrid viruses in nature. We noted a strong association between the CRESS DNA viruses and circular molecules identified in the water and browser organisms (*C. zealandicus*, *P. antipodarum* and *P. acuta*), and between water sediments and undefended prey species (*C. zealandicus*). However, we were unable to find any significant correlation of viral assemblages to the potential feeding relationships of the host aquatic invertebrates.

1. Introduction

Viruses are the most abundant biological entities on Earth (Edwards and Rohwer, 2005; Rohwer, 2003; Rohwer and Thurber, 2009). Besides their major ecological role in nutrient cycling, they are also frequently instrumental in regulating host population sizes, demographics, and spatial distributions at all trophic levels in the environment (Breitbart and Rohwer, 2005; Koonin et al., 2006; Suttle, 2007). However, viral dynamics and the 'flow' of viruses in ecosystems between different organisms are yet to be fully understood. Viral metagenomic methods have facilitated the discovery of numerous diverse circular replication-associated protein (Rep) encoding single-stranded (CRESS) DNA virus-like sequences, in a range of different environments; including sea water and marine organisms (Angly et al., 2006; Breitbart et al., 2015; Dayaram et al., 2013a,b, 2015a; Fahsbender et al., 2015; Labonte and Suttle, 2013; Ng et al., 2013; Rohwer and Thurber, 2009; Yoon et al., 2011), deep-sea vents (Yoshida et al., 2013), rice paddy soil (Kim et al., 2008), Antarctic lakes and ponds

* Corresponding author at: School of Biological Sciences, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand.
E-mail address: arvind.varsani@canterbury.ac.nz (A. Varsani).

http://dx.doi.org/10.1016/j.meegid.2016.02.011
1567-1348/

(Lopez-Bueno et al., 2009; Zawar-Reza et al., 2014), aquifers (Smith et al., 2013), reclaimed water and sewage (Kraberger et al., 2015; Ng et al., 2012; Rosario et al., 2009) and freshwater lakes (Hewson et al., 2013; Roux et al., 2012).

The diversity of potentially insect-transmissible CRESS DNA sequences has also been studied in various ecosystems through vector enabled metagenomics (Dayaram et al., 2013c, 2014, 2015b; Ng et al., 2011a, 2011b, 2013; Padilla-Rodriguez et al., 2013; Rosario et al., 2011, 2012a). This vector enabled metagenomic approach has been primarily employed to explore the role insects play in the transmission of plant-infecting geminiviruses, but has also revealed a high degree of insect-associated CRESS DNA diversity (Ng et al., 2011a, 2011b; Rosario et al., 2013, 2014, 2015).

The utility of viral metagenomics is not limited to the discovery of novel viruses and virus-like elements, it can also be a useful approach for studying the dynamics of viruses in various ecosystems. Understanding virus-host interactions in the context of an ecosystem can be difficult without access to baseline data on the viruses circulating in that ecosystem. Here we examine CRESS DNA virus diversity associated with different elements (water, benthic sediment, molluscs and insect larvae) within a freshwater lake ecosystem, and attempt to investigate associations between CRESS DNA viral assemblages within 1) aquatic invertebrates and the environments from which they were sampled, and 2) potential prey species and their predators.

2. Materials and methods

2.1. Sample collection

Samples were collected in December 2013 from Lake Sarah (43.0491°S, 171.7767°E; 577 m asl.), located near Cass in the South Island of New Zealand (Fig. 1). Aquatic invertebrates were collected from the shallow lake bottom and emergent vegetation using repeated 1-m D-net sweeps (1 mm mesh). Fingernail clams (*Musculium novazelandiae*; n = 46), New Zealand freshwater mussels (*Echyridella menzeisi*; n = 3), New Zealand mud snails (*Potamopyrgus antipodarum*; n = 26), tadpole snails (*Physella acuta*; n = 33), common midge larvae (*Chironomus zealandicus*; n = 27), yellow spotted dragonfly larvae (*Procordulia grayi*; n = 17), and red damselfly larvae (*Xanthocnemis zealandica*; n = 22) were sampled. These were sorted alive in the field and kept under refrigeration prior to processing. Samples of benthic sediment (40 ml, n = 1) and lake water (40 ml, n = 1) were also collected. Invertebrates were classified using identification keys (Winterbourn, 1973; Thorp and Covich, 1991; Winterbourn et al., 2006). All organisms were washed in sterile distilled water prior to processing for viral recovery.

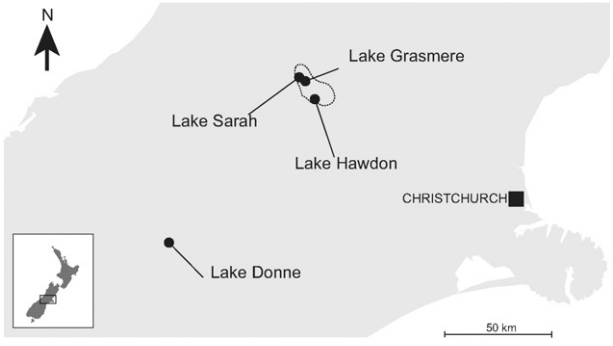

Fig. 1. Map of the partial region of the South Island of New Zealand showing Lake Sarah. The map also shows other nearby lakes where we have previously sampled Odonata larvae and identified various CRESS DNA viruses and DNA molecules (Dayaram et al., 2014). The dotted circle indicates the Cass Basin area.

2.2. Viral DNA extraction and enrichment of circular DNA

Samples were grouped according to species, and 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 per 5 g of tissue/sample, centrifuged (10,000 × *g* for 10 min) to pellet tissue and cellular debris, and then processed as previously described (Dayaram et al., 2013a,b, 2014, 2015a). In brief, the supernatant was passed through a 0.45 μm filter, followed by a 0.2 μm filter (Sartorius Stedim Biotech, Germany), viral DNA was recovered via the High Pure Viral Nucleic Acid kit (Roche, USA), and the circular DNA enriched using TempliPhi (GE Healthcare, USA).

2.3. Sequencing, de novo assembly and identification of CRESS DNA molecules

Prior to sequencing, the enriched circular DNA products of *P. grayi* and *X. zealandica* larvae were combined to form an Odonata/predator group, and the products of *P. antipodarum* and *P. acuta* combined to form a Gastropoda group. DNA from seven samples (water, sediment, Odonata, Gastropoda, *M. novazealandiae*, *E. menziesi* and *C. zealandicus*) was sequenced on an Illumina HiSeq 2000 (Illumina, USA) platform at the Beijing Genomics Institute (Hong Kong). Paired-end reads were de novo assembled using ABySS V1.5.0. (Simpson et al., 2009) as previously described (Dayaram et al., 2015a) and contigs >500 nts in length with similarities to previously identified eukaryotic CRESS DNA molecules were identified by BLASTx analysis (Altschul et al., 1990).

2.4. Recovery of CRESS DNA viral genomes

Contigs were selected for recovery based on the BLASTx analysis. To recover the complete circular DNA molecules of interest, an abutting pair of specific primers was designed for each contig (Supplementary Table 1) and used with Kapa HiFi HotStart polymerase (Kapa Biosystems, USA) for PCR. The resulting amplicons were gel purified, ligated into pJET1.2 (Thermo Fisher, USA), and Sanger sequenced at Macrogen Inc. (South Korea) by primer walking. The Sanger sequenced circular DNA viral sequence contigs were then assembled using DNA Baser Sequence Assembler (version 4.16; Heracle Biosoft S.R.L., Romania). An initial analysis of the circular DNA molecules was then carried out using BLASTx and tBLASTx (Altschul et al., 1990), and was followed by annotation of the genomes with key features, including the major open reading frames (ORFs), conserved motifs, and stem-loop structures.

2.5. Classification and putative species assignment

We grouped the CRESS DNA viruses and circular DNA molecules based on the classification system (type groupings I–VIII) proposed by (Rosario et al., 2012b) (Fig. 2). The Reps of the CRESS DNA viruses and molecules recovered in this study were aligned using PROMALS3D (Pei et al., 2008), and this alignment was used to infer a maximum-likelihood phylogenetic tree using PhyML (Guindon et al., 2010). This included a VT + I + G + F substitution model selected using ProtTest (Darriba et al., 2011), as well as SH-like branch support. Branches with less than 80% support were collapsed. We tentatively defined CRESS DNA species as groups of CRESS DNA elements that 1) formed well supported (SH-like branch support p-value > 0.9) monophyletic clusters in the Rep phylogenetic tree, and 2) shared greater than 75% Rep amino acid sequence identity. Individual CRESS DNA elements with predicted Rep amino acid sequences with <75% identity to all other known CRESS DNA elements were assigned their own species (Fig. 3A). All pairwise nucleotide and amino acid identities were calculated using SDT v1.2 (Muhire et al., 2014) with MUSCLE option (Edgar, 2004).