Virus Research 177 (2013) 209–216

Contents lists available at ScienceDirect

Virus Research

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

Short communication

Novel myco-like DNA viruses discovered in the faecal matter of various animals

Alyssa Sikorski[a], Melanie Massaro[b], Simona Kraberger[a], Laura M. Young[a], Darren Smalley[a], Darren P. Martin[c], Arvind Varsani[a,d,e,*]

[a] School of Biological Sciences, University of Canterbury, Christchurch 8140, New Zealand
[b] School of Environmental Sciences, Charles Sturt University, Albury, NSW, Australia
[c] Computational Biology Group, Institute of Infectious Diseases and Molecular Medicine, University of Cape Town, Cape Town 7925, South Africa
[d] Electron Microscope Unit, Division of Medical Biochemistry, Department of Clinical Laboratory Sciences, University of Cape Town, Observatory 7700, South Africa
[e] Biomolecular Interaction Centre, University of Canterbury, Christchurch 8140, New Zealand

ARTICLE INFO

Article history:
Received 16 July 2013
Received in revised form 14 August 2013
Accepted 20 August 2013
Available online 30 August 2013

Keywords:
ssDNA viruses
Gemini-like viruses
Geminiviruses
Fungal DNA viruses
Faeces
SsHADV-1

ABSTRACT

A wide variety of novel single-stranded DNA (ssDNA) viruses have been found in faecal matter of chimpanzees, cows, rodents, bats, badgers, foxes and pigs over the last few years. Using a combination of rolling circle amplification coupled with restriction enzyme digests based approach as well as a next generation sequencing informed approach, we have recovered fourteen full genomes of ssDNA viruses which exhibit genomic features described for members of the recently proposed gemycircularvirus group from a wide variety of mammal and bird faecal samples across New Zealand. The fourteen novel ssDNA viruses (2122–2290 nt) encode two major proteins, a replication associated protein (Rep) and a capsid protein (Cp) which are bi-directionally transcribed. Interestingly, the Rep of these novel viruses are similar to gemycircularviruses detected in insects, cassava leaves, and badger faecal matter, the novel viruses share sequence similarities with the mycovirus sclerotinia sclerotiorum hypovirulence-associated DNA virus 1 (SsHADV-1) and Rep-like sequences found in fungal genomes. Pairwise sequence similarities between the 14 novel genomes with other related viral isolates (gemycircularviruses) indicated that they share greater than 55.8% genome-wide identity. Additionally, they share between 55% and 59% pairwise identity with putative novel ssDNA virus genomes recently isolated from sewage baminivirus, niminivirus and nephavirus. Based on the similarities to SsHADV-1 and Rep-like sequences found in fungal genomes, these novel gemycircularviruses may infect fungi.

Animal faecal matter can contain a range of diverse viruses including those infecting the source of the faeces, the animals and/or plants they eat, and their commensal gut microbes (Breitbart et al., 2003, 2008; Fukutomi et al., 1996; Gaffney et al., 2012; Nazir et al., 2011; Pedersen et al., 2009; Theil and McCloskey, 1995; Zhang et al., 2006). To some extent, viruses sampled within faeces can be used to explore viral diversity at the ecosystem scale and therefore present a convenient means of non-invasively monitoring the environment for the presence of either viruses with known socio-economic impacts or viruses with the potential to emerge as pathogens in humans or domesticated animals and plants. Among the most interesting viruses found so far within animal faecal samples has been an astonishing assortment of highly divergent single-stranded DNA (ssDNA) viruses. Numerous viruses that will likely be assigned to completely new ssDNA virus families and genera have been found within the faeces of chimpanzees (Blinkova et al., 2010; Li et al., 2010a), cows (Kim et al., 2012), rodents (Phan et al., 2011), bats (Ge et al., 2011; Li et al., 2010b), badgers (van den Brand et al., 2012), foxes (Bodewes et al., 2013), New Zealand fur seal (Sikorski et al., 2013b) and pigs (Sachsenröder et al., 2012; Shan et al., 2011; Sikorski et al., 2013a). Crucial to the success of sampling diverse viruses from faeces has been the development of techniques for both the efficient purification of viral particles and the non-specific amplification of sequences and/or identification of viral genomic sequences (Breitbart et al., 2003, 2008; Rosario et al., 2009; Sikorski et al., 2013a,c; Thurber et al., 2009; Zhang et al., 2006).

We applied a protocol used previously for the isolation of ssDNA viruses from porcine seal faecal matter (Sikorski et al., 2013a,b) and parrot nesting material (Sikorski et al., 2013c) to investigate ssDNA viruses in faecal samples of domesticated and wild

* 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 addresses: arvind.varsani@canterbury.ac.nz, avarsani@gmail.com (A. Varsani).

0168-1702/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.virusres.2013.08.008

mammals and birds from across New Zealand (Supplementary Table 1). We specifically focus here on the detailed analysis of viruses discovered within these samples that had detectable sequence similarities with the recently discovered geminivirus-like ssDNA mycovirus virus sclerotinia sclerotiorum hypovirulence-associated DNA virus 1 (SsHADV-1) (Yu et al., 2010).

SsHADV-1 was the first fungal-infecting circular ssDNA virus to be identified and was found to have icosahedral particles encapsidating a 2166 nt genome (Yu et al., 2010). Further, recent studies by Yu et al. (2013) demonstrated that *S. sclerotiorum* is susceptible to infection by extracellular inoculation of SsHADV-1 virions and that the virions are stable for extended periods of time, however, they probably have a limited host range (currently, known range extending to *Sclerotinia minor* and *Sclerotinia nivali*). The SsHADV-1 genome encodes a replication associated protein (Rep), that while clearly shares homology to those encoded by a range of other diverse ssDNA viruses, is most similar to those encoded by geminiviruses and actually even contains a geminivirus-like Rep sequence motif (GRS) (Nash et al., 2011; Yu et al., 2010). Infectivity assays carried out by Yu et al. (2013) clearly demonstrate that SsHADV-1 was unable to infect and move in plant cells. SsHADV-1 has recently been discovered in benthic and bank river sediments from two urban rives in New Zealand (Kraberger et al., 2013) suggesting that SsHADV-1 may be more common in ecosystems than previously thought.

Since 2011, a number of other apparent ssDNA viral genomes with high degrees of similarity and similar genomic architecture to that of SsHADV-1 have been recovered from the leaves of a Ghanaian cassava plant [cassava associated circular DNA virus (CasCV; Dayaram et al., 2012)], from the faeces of a European badger in the Netherlands [Meles meles faecal virus (MmFV; van den Brand et al., 2012)], from mosquitoes in the USA [mosquito VEM SDBVL-G virus (MvemV); (Ng et al., 2011)] and from dragonflies in the Kingdom of Tonga and the USA [dragonfly-associated circular viruses-1, 2, and 3 (DfasCV-1, -2 and -3; (Rosario et al., 2012a)]. Together with SsHADV-1, these viral isolates represent a novel group of viruses and the name gemycircularviruses (Gemini-like myco circular viruses) has been proposed by Rosario et al. (2012a). It is also apparent that sequences with high degrees of similarity to the Rep encoding genes of these viruses are integrated within the genomes of various fungi including *Aspergillus fumigatus*, *Colletotrichum higginsianum*, *Laccaria bicolor*, *Magnaporthe oryzae*, *Nectria haematococca*, *Serpula lacrymans* var *lacrymans* and *Tuber melanosporum* (Liu et al., 2011). Based on similarities to the mycovirus SsHADV-1 and close phylogenetic association with rep-like sequences in fungal genomes, it is highly likely that gemycircularviruses infect fungi, etc.

All faecal samples analysed here were collected and stored at −20 °C prior to use. Avian faecal samples collected on the Chatham Islands, New Zealand, were grouped by bird species per annual field season (for example, all Chatham Island black robin (*Petroica traversi*) samples collected in the 3 month field season of 2008 were grouped into a single sample). Using a disposable spatula, five grams of faecal matter was added to 5 ml of SM buffer (0.1 M NaCl, 50 mM Tris/HCl – pH 7.4, 10 mM $MgSO_4$), homogenised and centrifuged (10,000 × *g* for 10 min on a bench top microfuge). The supernatant was filtered through a 0.45 μm followed by a 0.2 μm pore size syringe filter (Sartorius Stedim Biotech, Germany) and 200 μl of the sample filtrate was used for viral isolation using the High Pure Viral Nucleic Acid Kit (Roche, USA) according to the manufacturer's instructions. To enrich for circular single-stranded DNA, 1 μl of purified viral nucleic acid was subjected to rolling circle amplification (RCA) using the Illustra™ TempliPhi Kit (GE Healthcare, USA). The RCA products were cut in different reactions using a variety of restriction endonucleases (*Bam*HI, *Kpn*I, *Xmn*I, *Eco*RI, and *Eco*RV). Fragments in the size range of 1.5–6 kb were ligated to pGEM3Zf (+) (Promega, USA) and pUC19 (Thermo Scientific, USA) plasmid cloning vector that had been restricted with the appropriate restriction enzyme. The resulting cloned amplicons were sequenced by primer walking at Macrogen Inc. (Korea).

Additionally, 1 μl of viral nucleic acid from all the samples was combined and enriched by RCA for circular ssDNA. Linear concatemers of circular DNA amplicons were sequenced at Beijing Genomics Institute (Hong Kong) using an Illumina HiSeq 2000 (Illumina, USA) platform. The resulting paired end reads were assembled (Kmer = 64) using Abyss v 1.3.5 (Simpson et al., 2009). Assembled contigs that were >1000 bases were screened for homology to encoded proteins of available sequences using tBLASTx (Altschul et al., 1990) implemented in KoriBlast v 3.4 (KoriLog Bioinformatics Solution, France). We found 14 unique contigs with high degrees of similarity to SsHADV-1-like Rep sequences.

Based on the resulting sequence data from the RCA/restriction enzyme digests and assembled Illumina paired end sequencing reads, we designed back-to-back (abutting) primers based on the conserved region of the Rep sequences (Table 1) to recover all the full length viral regomes. These primers were used to PCR screen all the RCA reactions from all the faecal samples and full length viral genomes were recovered using Kapa HiFi HotStart DNA polymerase (Kapa Biosystems, USA). The resulting amplicons were ligated to pJET 1.2 vector (Thermo Scientific, USA) and sequenced at Macrogen Inc. (Korea) by primer walking. Full genomes were assembled using DNAMAN (version 7; Lynnon Biosoft, Canada).

Using these two approaches a total of fourteen distinct SsHADV-like ssDNA viruses were recovered from eleven faecal samples (Tables 1 and 2; GenBank accession numbers: KF371630–KF371643) representing ostrich (*Struthio camelus*; *n* = 1; 2011), blackbird (*Turdus merula*; *n* = 2; 2009 and 2011), black robin (*Petroica traversi*; *n* = 2; 2011); Chatham Island warbler (*Gerygone albofrontata*; *n* = 3; 2011), Mallard duck (*Anas platyrhynchos*; *n* = 1; 2012), European starling (*Sturnus vulgaris n* = 1; 2009), pig (*Sus scrofa*; *n* = 1; 2011) rabbit (*Oryctolagus cuniculus*; *n* = 1; 2009), New Zealand fur seal (*Arctocephalus forsteri*; *n* = 1; 2012) and sheep (*Ovis aries*; *n* = 1; 2009). Six were recovered using the first technique employing a RCA-restriction enzyme digest based approach followed by sequencing and inverse PCR using sequence specific primer pairs to amplify full genomes and eight using the second technique employing back-to-back sequence specific primers designed based on the assembled Illumina sequencing data.

The 14 viral isolates exhibit all the genomic features identified in gemycicularviruses including an intergenic region with: (1) a nonanucleotide motif (Table 2) nested within 4–12 nucleotide long inverted repeat sequence with the capacity to form a hairpin structure analogous to those found at the origin of virion strand replication (*v-ori*) in geminiviruses, circoviruses, and nanoviruses (Table 2) (Rosario et al., 2012b); (2) an open reading frame in the virion sense likely encoding a capsid protein; and (3) two open reading frames (other than the isolate from the ostrich faecal sample) in the complementary sense that, following splicing of a complementary strand transcript likely expresses Rep. Such intron containing *rep* genes are similar to those found in certain geminiviruses, such as those in of the genus mastrevirus (Dekker et al., 1991; Gutierrez, 1999; Mullineaux et al., 1990; Schalk et al., 1989; Wright et al., 1997) and becurtoviruses (beet curtly top Iran virus) (Heydarnejad et al., 2013). While we identified potential intron acceptor and donor sites that would yield a full length *rep* transcript upon splicing in all of the isolates, the possibility cannot be excluded that, as is known to occur in mastreviruses, a variant of Rep (RepA) might be expressed which has a N-terminus that is identical to that of Rep and a C-terminus that is unique.

There is significant conservation in the unique rolling circle replication (RCR) motifs (I, II and III), the GRS, and helicase motifs (Table 2). For all gemycircularviruses described to date, including