Virology 504 (2017) 114–121

Contents lists available at ScienceDirect

Virology

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

Evolutionary history of ssDNA bacilladnaviruses features horizontal acquisition of the capsid gene from ssRNA nodaviruses

Darius Kazlauskas[a], Anisha Dayaram[b,c], Simona Kraberger[b,d], Sharyn Goldstien[b], Arvind Varsani[b,e,f,*], Mart Krupovic[g,**]

[a] Institute of Biotechnology, Vilnius University, Saulėtekio av. 7, Vilnius 10257, Lithuania
[b] School of Biological Sciences, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand
[c] Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, CO 80523, USA
[d] Department of Wildlife Diseases, Leibniz Institute for Zoo and Wildlife Research, Berlin 10315, Germany
[e] The Biodesign Center for Fundamental and Applied Microbiomics, School of Life sciences, Center for Evolution and Medicine, Arizona State University, Tempe, AZ 85287, USA
[f] Structural Biology Research Unit, Department of Clinical Laboratory Sciences, University of Cape Town, Observatory 7700, South Africa
[g] Unité Biologie Moléculaire du Gène chez les Extrêmophiles, Department of Microbiology, Institut Pasteur, 25 rue du Docteur Roux, Paris 75015, France

ARTICLE INFO

Keywords:
ssDNA viruses
CRESS DNA viruses
Bacilladnavirus
Capsid protein
Nodaviridae

ABSTRACT

Bacilladnaviruses have single-stranded (ss) DNA genomes and infect diatoms, a major group of unicellular algae widespread in aquatic habitats. Despite their ecological importance, the provenance and relationships of bacilladnaviruses to other eukaryotic viruses remain unclear. Accordingly, they are currently classified into the 'floating' genus *Bacilladnavirus*. Here we present three new bacilladnavirus genomes recovered from a mollusc *Amphibola crenata* and benthic sediments from the Avon-Heathcote estuary in New Zealand. Our analysis shows that the rolling-circle replication-initiation proteins of bacilladnaviruses display unique conserved motifs and in phylogenetic trees form a monophyletic clade separated from other groups of ssDNA viruses. Unexpectedly, distant homology detection combined with structural modeling indicates that bacilladnavirus capsid proteins are homologous to those of ssRNA viruses from the *Nodaviridae* family. Considering the sequence diversity within the expanding *Bacilladnavirus* genus, we argue that classification of these viruses has to be revised and the current genus upgraded to the family level.

1. Introduction

Viruses with single-stranded (ss) DNA genomes are among the smallest known viruses. With few exceptions (Mochizuki et al., 2012; Pietilä et al., 2016), they have less than 10 genes but many suffice with just 2 genes, one for the capsid protein (CP) and the other for a genome replication initiation protein (Krupovic, 2013; Rosario et al., 2012b). Despite such minimalistic genetic layout, ssDNA viruses are abundant and widespread in diverse environments, indicating that they are competitive members of the global virome. ssDNA viruses infecting eukaryotes are particularly diverse. These are currently classified by the International Committee on Taxonomy of Viruses (ICTV) into 7 families and one unassigned ('floating') genus (Krupovic et al., 2016). Members of the families *Anelloviridae*, *Circoviridae*, *Parvoviridae* and *Bidnaviridae* infect animals; those of the *Geminiviridae* and *Nanoviridae* infect plants; whereas certain viruses of the recently introduced family *Genomoviridae* can infect both phytopathogenic fungi and mycophagous insects (Krupovic et al., 2016; Liu et al., 2016). Finally, the unassigned genus *Bacilladnavirus* includes viruses that infect diatoms (Nagasaki, 2008), a major group of unicellular algae (class Bacillariophyceae), with as many as 200,000 extant species found in aquatic environments (Mann and Droop, 1996).

The rapid evolution and diversification of ssDNA viruses is fueled by very high mutation and recombination rates in their small genomes (Duffy and Holmes, 2009; Firth et al., 2009; Grigoras et al., 2010; Harkins et al., 2009; Lefeuvre et al., 2009; Martin et al., 2011; Roux et al., 2013; Streck et al., 2011). Interestingly, recombination occurs not only between closely related viruses but also among evolutionarily unrelated viruses. Indeed, several cases of gene transfer between viruses with RNA and DNA genomes have been reported (Diemer

* Corresponding author at: The Biodesign Center for Fundamental and Applied Microbiomics, School of Life sciences, Center for Evolution and Medicine, Arizona State University, Tempe, AZ 85287, USA.
** Corresponding author at: Unité Biologie Moléculaire du Gène chez les Extrêmophiles, Department of Microbiology, Institut Pasteur, 25 rue du Docteur Roux, Paris 75015, France.
E-mail addresses: d.kazlauskas@ibt.lt (D. Kazlauskas), anisha.dayaram@gmail.com (A. Dayaram), simona.kraberger@gmail.com (S. Kraberger), sharyn.goldstien@canterbury.ac.nz (S. Goldstien), arvind.varsani@asu.edu (A. Varsani), krupovic@pasteur.fr (M. Krupovic).

http://dx.doi.org/10.1016/j.virol.2017.02.001
Received 12 January 2017; Received in revised form 1 February 2017; Accepted 2 February 2017

and Stedman, 2012; Krupovic and Koonin, 2014; Krupovic et al., 2009). For instance, viruses of the proposed family 'Cruciviridae' (Quaiser et al., 2016) have apparently emerged following a recombination event between an ssDNA virus which donated a gene for the rolling-circle replication initiation protein (Rep) and an ssRNA virus which provided a gene for the tombusvirus-like CP (reviewed in Krupovic (2013) and Stedman (2015)). Generally, CPs from one family of ssDNA viruses are not recognizably similar to the corresponding proteins from another family.

Contrasting the high diversity of CPs, proteins responsible for viral genome replication are considerably more conserved. With the notable exception of bidnaviruses and parvoviruses, eukaryotic ssDNA viruses have circular genomes that are replicated by the rolling-circle mechanism initiated by the virus-encoded endonuclease, Rep (Krupovic, 2013; Rosario et al., 2012b). The Reps from different virus families display conserved domain organization with the N-terminal endonuclease domain and the C-terminal superfamily 3 helicase domain (Krupovic, 2013; Gorbalenya et al., 1990). Notably, such domain organization is not found in characterized viruses of bacteria and archaea, but is conserved in a great variety of uncultured and unclassified ssDNA viruses discovered by metagenomics approaches (Simmonds et al., 2017). Conservation of the Rep proteins across several different assemblages of ssDNA viruses led to their informal unification within a supergroup referred to as the CRESS DNA viruses (for *circular Rep-encoding ssDNA viruses*) (Rosario et al., 2012a). Phylogenetic analysis of the eukaryotic ssDNA virus Reps confirmed the monophyly of established families and shed light on the evolutionary relationships between the major virus groups. For instance, in Rep-based phylogenetic trees, genomoviruses form a sister group to geminiviruses (Krupovic et al., 2016), whereas Reps of nanoviruses are closely related to those of alphasatellites (Simmonds et al., 2017). It has been also suggested that Reps of nanoviruses and circoviruses are closer to each other than they are to Reps of geminiviruses (Gibbs and Weiller, 1999). By contrast, the relationship of bacilladnaviruses to other eukaryotic ssDNA viruses remains largely unresolved.

Chaetoceros salsugineum DNA virus 01 (CsalDNAV; AB193315) is currently the only officially classified member of the *Bacilladnavirus* genus (Nagasaki et al., 2005). However, several potential members of the genus infecting different diatom species have been isolated and their genomes completely sequenced. These include Chaetoceros sp. DNA virus 7 (Csp07DNAV; AB844272) infecting *Chaetoceros* sp. strain SS628-11 (Kimura and Tomaru, 2013), Chaetoceros lorenzianus DNA virus (ClorDNAV; AB553581) infecting *Ch. lorenzianus* (Tomaru et al., 2011b), Chaetoceros setoensis DNA virus (CsetDNAV; AB781089) of *Ch. setoensis* (Tomaru et al., 2013), and Chaetoceros tenuissimus DNA viruses (CtenDNAV; AB597949) type I (Tomaru et al., 2011a) and type II (AB971658 – AB971660) (Kimura and Tomaru, 2015) infecting *Ch. tenuissimus*.. Partial genome sequences are available for Chaetoceros virus YT-2008 (Csp05DNAV; AB647334) of *Chaetoceros* sp. strain TG07-C28 (Toyoda et al., 2012), Chaetoceros debilis DNA virus (CdebDNAV; AB504376) infecting *Chaetoceros debilis* (Tomaru et al., 2008) and Thalassionema nitzschioides DNA virus (TnitDNAV; AB781284) infecting *Thalassionema nitzschioides* (Tomaru et al., 2012). In addition, a putative bacilladnavirus genome, Bacillariodnavirus LDMD-2013 (KF133809), has been assembled from metagenomic sequences (McDaniel et al., 2014). Characterized bacilladnaviruses carry circular ssDNA genomes of ~5.5–6 kb encapsidated within isometric particles of 33–38 nm in diameter (Kimura and Tomaru, 2015). Interestingly, it has been demonstrated that digestion of the bacilladnavirus genomes with S1 nuclease results in short linear dsDNA fragments of variable sizes (< 1 kb), suggesting that the genomes might be partially double-stranded. The number and location of such fragments within the viral genomes varies (Kimura and Tomaru, 2015; Tomaru et al., 2013), whereas their function, if any, remains unresolved. Despite the potentially profound role of bacilladnaviruses in aquatic ecosystems, it is unclear whether the 'floating' genus *Bacilladnavirus* should be included into any of the existing virus families or whether it is sufficiently distinct to be classified into a separate family.

Here, we present three new bacilladnavirus genomes recovered from a gastropod mollusc *Amphibola crenata* and benthic sediments collected in the Avon-Heathcote estuary (New Zealand). Phylogenetic analysis of Rep proteins showed that bacilladnaviruses form a distinct branch among CRESS DNA viruses. Consistently, sequence analysis revealed unique motifs within the bacilladnavirus Reps, indicating that they are not closely related to any of the existing families of ssDNA viruses. Finally, distant homology detection analysis and subsequent homology modeling indicate that bacilladnavirus CPs are homologous to the CPs of ssRNA viruses of the *Nodaviridae* family, suggesting a horizontal gene transfer event between these ssDNA and ssRNA viruses. Our results provide important information on the diversity, distribution and evolution of bacilladnaviruses. We also conclude that taxonomic re-evaluation of the *Bacilladnavirus* genus is in order.

2. Results

2.1. *New members of the genus* Bacilladnavirus

Three circular virus genomes were *de novo* assembled from metagenomic reads generated from samples recovered from a gastropod mollusc *Amphibola crenata* and benthic sediments of the Avon-Heathcote Estuary in New Zealand. The validity of the contigs was confirmed by inverse PCR, cloning and Sanger sequencing (see Materials and Methods). All three genomes shared significant sequence similarity with bacilladnaviruses (see below). Thus, the two genomes recovered from *A. crenata* were denoted Amphibola crenata associated bacilladnavirus 1 (AcrBV1) and AcrBV2, whereas the one from the estuary sample was called Avon-Heathcote Estuary associated bacilladnavirus (AHEaBV). Genomes of AcrBV1, AcrBV2 and AHEaBV are 4729, 4576 and 4742 nt in length, respectively, and are the smallest bacilladnavirus genomes characterized thus far.

Similar to other bacilladnaviruses, the new genomes each contain four open reading frames (ORFs 1–4) (Fig. 1). We used sensitive sequence searches to predict their functions (Table S1). Two of the detected viral ORFs encode for putative capsid (ORF3) and replication (ORF4) proteins, whereas the other two (ORF1 and ORF2) show no significant similarity to sequences in public databases (Table S1), although they are conserved in the two moderately similar viruses AcrBV1 and AcrBV2 (Fig. 1). Notably, despite the lack of sequence similarity between the corresponding ORF1 and 2, the genomes of all analyzed bacilladnaviruses are collinear, with 2 ORFs encoded on the sense strand and the other two on the complementary one.

BLASTx analysis of the Rep proteins of AcrBV1 and AcrBV2 shows that they are most similar to each other (58% identity over 504 aa), whereas the next closest hit is to the corresponding protein of Bacillariodnavirus LDMD-2013 (54% identity over 418 aa), an uncultured virus genome originating from a marine sample collected in Tampa Bay, Florida, USA (McDaniel et al., 2014). By contrast, the Rep of AHEaBV is more divergent and most similar to the protein of Chaetoceros virus YT-2008 (43% identity over 409 aa). The CPs of AcrBV1, AcrBV2 and AHEaBV are most similar to the CP of Bacillariodnavirus LDMD-2013 (Table S1). Notably, CP homologs from all bacilladnaviruses were recovered in the first PSI-BLAST (Altschul et al., 1997) iteration and subsequent iterations did not yield any additional CPs from other viruses.

2.2. *Bacilladnaviruses encode nodavirus-like capsid proteins*

To gain insights into the fold and provenance of the bacilladnaviral CP, we performed sensitive profile-profile comparisons using HHpred server (Söding et al., 2005). Unexpectedly, this analysis showed that bacilladnavirus CPs display moderate similarity (with average prob-