Virus Research 164 (2012) 114–121

Contents lists available at SciVerse ScienceDirect

Virus Research

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

Review

Rapidly expanding genetic diversity and host range of the *Circoviridae* viral family and other Rep encoding small circular ssDNA genomes

Eric Delwart[a,b,*], Linlin Li[a,b]

[a] *Blood Systems Research Institute, San Francisco, CA, USA*
[b] *Department of Laboratory Medicine, University of California, San Francisco, CA, USA*

ARTICLE INFO

Article history:
Available online 6 December 2011

Keywords:
Circovirus
Cyclovirus
Circoviridae
Rep protein
Deep sequencing
Circular ssDNA genome

ABSTRACT

The genomes of numerous circoviruses and distantly related circular ssDNA viruses encoding a rolling circle replication initiator protein (Rep) have been characterized from the tissues of mammals, fish, insects, plants (geminivirus and nanovirus), in human and animal feces, in an algae cell, and in diverse environmental samples. We review the genome organization, phylogenetic relationships and initial prevalence studies of cycloviruses, a proposed new genus in the *Circoviridae* family. Viral fossil *rep* sequences were also recently identified integrated on the chromosomes of mammals, frogs, lancelets, crustaceans, mites, gastropods, roundworms, placozoans, hydrozoans, protozoans, land plants, fungi, algae, and phytoplasma bacterias and their plasmids, reflecting the very wide past host range *of rep bearing viruses*. An ancient origin for viruses with *Rep*-encoding small circular ssDNA genomes, predating the diversification of eukaryotes, is discussed. The cellular hosts and pathogenicity of many recently described *rep*-containing circular ssDNA genomes remain to be determined. Future studies of the virome of single cell and multicellular eukaryotes are likely to further extend the known diversity and host-range of small *rep*-containing circular ssDNA viral genomes.

Contents

1. Introduction

Members of the genus *Circovirus* in the family *Circoviridae*, are non-enveloped, icosahedral viruses with a single-stranded circular DNA (ssDNA) genome of approximately 2 kb, the smallest known autonomously replicating viral genomes (Todd et al., 2005). Circoviruses infect numerous bird species including parrots, pigeons, gulls, ducks, geese, swans, ravens, canaries, finches, starlings, and chickens (Niagro et al., 1998; Mankertz et al., 2000; Todd et al., 2001a, 2007; Hattermann et al., 2003; Johne et al., 2006; Stewart et al., 2006; Halami et al., 2008; Li et al., 2011). To date only two circoviruses have been extensively documented to replicate in a mammal, Porcine circovirus 1 and 2 (PCV1 and PCV2) (Allan and

* Corresponding author at: Blood Systems Research Institute, San Francisco, CA, USA.
E-mail address: delwarte@medicine.ucsf.edu (E. Delwart).

0168-1702/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.virusres.2011.11.021

Ellis, 2000; Mankertz et al., 2004). PCV1 is generally considered non-pathogenic while PCV2 infection can be either asymptomatic or cause a variety of clinical symptoms with significant economic impact (Finsterbusch and Mankertz, 2009; Todd et al., 2001b; Segales et al., 2005; Chae, 2005; Opriessnig et al., 2007; Grau-Roma et al., 2011).

Circoviruses have an ambisense genome organization containing two major inversely arranged open reading frames, encoding the rolling circle replication initiator protein gene (*rep*) and capsid protein gene (*cap*) (Todd et al., 2005). A stem-loop structure with a conserved 9 bases motif in the loop, located between the 5′-ends of the two main ORFs, is required to initiate the replication of the viral genome. The replication complex consists of Rep and a shorter Rep' protein with a different carboxy termini derived from a spliced transcript. Following cell infection, a double stranded template genome is first generated by cellular DNA polymerase 1 extending a small RNA primer. Rep and Rep' bind the stem loop, cutting a nick in the plus strand and a host-encoded DNA polymerase then extends the 3′ hydroxyl to copy the complementary circle using a rolling circle replication mechanism (Steinfeldt et al., 2001, 2007; Faurez et al., 2009). The rolling circle replication strategy of PCV is similar to that of plant Geminivirus and Nanovirus and of bacterial plasmids in the pT181 family (Timchenko et al., 1999; del Solar et al., 1998; Gutierrez, 1999).

There has been a recent surge of small circular DNA genomes containing a *rep* gene discovered from different sources using different methods. In vitro rolling circle amplification (Haible et al., 2006; Rector et al., 2004), high-throughput sequencing (Blinkova et al., 2009; Rosario et al., 2009a; Ng et al., 2011a; Li et al., 2010a) and/or degenerate/consensus PCR have all been extensively used to identify novel *rep* containing circular DNA genomes in tissues (Li et al., 2010b, 2011) and feces of mammals (Ge et al., 2011; Li et al., 2010a), fish (Lorincz et al., 2011), insects (Ng et al., 2011b; Rosario et al., 2011) and in environmental samples (Rosario et al., 2009a,b; Blinkova et al., 2009; López-Bueno et al., 2009; Kim et al., 2008). Our understanding of the extensive genetic diversity of the *Circoviridae* and of distantly related viral families of *rep* bearing small circular ssDNA genome has therefore rapidly increased.

2. Ancient origin of small single stranded circular DNA genomes encoding Rep

Multiple lines of evidence point to an ancient origin for circoviruses and related genomes. The recent detection of genetically decayed fossil circovirus-like sequences integrated into the chromosomes of various mammals (as well as a frog) yielded estimates that these viruses replicated in mammals at least and possibly more than 100 million years ago (Katzourakis and Gifford, 2010; Belyi et al., 2010). The detection of Rep-encoding ORFs integrated in the chromosomes of the common parasitic protozoa *Giardia duodenalis* and *Entamoeba histolytica* may also reflect past replication of related genomes and integration in those hosts possibly due to the DNA binding, cutting and ligating Rep activity normally used for rolling circle replication (Gibbs et al., 2006). A geminivirus-like genome was similarly observed in the genome of tobacco plants (Bejarano et al., 1996). A recent comprehensive search of NCBI databases has also greatly increased the number of eukaryotic genomes known to contain viral *rep*-like genes (Liu et al., 2011).

The viral *rep* gene may have originated through recombination between unrelated viruses, with its N-terminal region being most related to that encoded by small single-stranded circular DNA viruses with segmented genomes of the *Nanoviridae* family (infecting plants) and a C-terminal region related to the RNA-binding 2C helicase protein of positive strand RNA picorna-like viruses (Gibbs and Weiller, 1999). The similar rolling circle replication strategy employed by the *Geminiviridae* and *Nanoviridae* families infecting plants, the *Circoviridae* infecting mammals, birds and possibly fishes and insects, and of some bacterial plasmids, may also reflect a common evolutionary history for these genomes (Faurez et al., 2009). Sequence similarities between the Rep of some plasmids and those of geminiviruses and parvoviruses (which also replicate via a rolling circle DNA replication mechanism) led to the hypothesis that these eukaryotic viruses evolved from eubacterial replicons (Koonin and Ilyina, 1992; Ilyina and Koonin, 1992). The phylogenetic closeness of the *rep* of plant Geminiviruses to the *rep* of plasmids of wall-less, plant infecting, phytobacteria led to the proposal that a phytoplasma plasmid may have evolved into a Geminivirus virus (Krupovic et al., 2009). While no sequence similarity was detected between geminivirus capsids and other proteins, their predicted structure was most like that that of a capsid protein from a ssRNA plant virus (satellite tobacco necrosis virus), a helper virus-dependent genome encoding only a capsid gene (Krupovic et al., 2009). Geminiviruses may have therefore captured their capsid gene from a plant RNA virus (Faurez et al., 2009). An alternative theory has been proposed that phytoplasmal plasmids acquired their *rep* by horizontal transfer from a geminivirus (Saccardo et al., 2011). The $T = 1$ icosahedral structure of geminiviruses (Böttcher et al., 2004; Zhang et al., 2001) and circoviruses (Khayat et al., 2011) both consist of capsids with a canonical viral jelly roll structure (eight stranded beta-barrel fold) which may also reflect a common evolutionary path. Tandem repeats of either Geminivirus or porcine circovirus encoding intact *rep* genes and their stem loop origins of replication can generate replicative forms in bacterial hosts that are indistinguishable from those in their eukaryotic hosts (Rigden et al., 1996; Cheung, 2006; Gibbs and Weiller, 1999; Selth et al., 2002). Collectively these observations support the hypothesis that small circular ssDNA viral genomes originated from prokaryotic episomal replicons (Koonin and Ilyina, 1992; Ilyina and Koonin, 1992).

3. Cyclovirus, a new genus in the *Circoviridae* family

As part of a metagenomics based search for new viruses, the viral nucleic acids in the feces of children from developing countries were randomly amplified and sequenced (Victoria et al., 2009). The initial cyclovirus genome fragment identified encoded a partial Rep protein detected through BLASTx sequence similarity searches against all viral protein sequences (Victoria et al., 2009). Given that circoviruses have small circular DNA genomes, the rest of the viral genome was then amplified by inverse PCR (Li et al., 2010b).

Like circoviruses, the cycloviruses have small circular ambisense DNA genomes of 1.7–1.9 kb (Table 1) containing two major inversely arranged ORFs encoding the putative Rep and Cap proteins. The rep of one genome (CyCV-TN25) was interrupted by a small putative 171 bases intron (Li et al., 2010a). Relative to circoviruses, the Rep and Cap proteins of cycloviruses are slightly shorter and the 3′-intergenic regions between the stop codons of the two major ORFs were either absent or consisted of only a few bases (Li et al., 2010b, 2011). The 5′-intergenic regions between the start codons of the *rep* and *cap* ORFs of cycloviruses were relatively larger than those of circoviruses and also contained a highly conserved stem-loop structure with a distinct loop nonamer sequence (Table 1). In the Rep N-terminus half several motifs associated with rolling circle replication (FTLNN, TPHLQG and YCSK) and dNTP-binding (GXGSK) were identified, with some alterations. Conserved amino acid motifs associated with 2C helicase function of some picorna-like viruses were also identified in the carboxy half of Rep (WWDGY, DDFYGW, and DRYP). The N-terminal region of the cyclovirus Cap proteins was highly basic and arginine-rich, as is typical for circoviruses capsid proteins. Cycloviruses are therefore distinguishable from circoviruses based on several unique genome