Journal of Invertebrate Pathology 140 (2016) 83–96

Contents lists available at ScienceDirect

Journal of Invertebrate Pathology

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

Review article

# Diversity of small, single-stranded DNA viruses of invertebrates and their chaotic evolutionary past

Peter Tijssen [a,*], Judit J. Pénzes [a], Qian Yu [a], Hanh T. Pham [a], Max Bergoin [a,b,*]

[a] Laboratoire de Virologie (Bldg 18), Institut National de Recherche Scientifique-Institut Armand-Frappier, 531 Boul. des Prairies, Laval, QC, H7V 1B7, Canada
[b] Laboratoire de Pathologie Comparée, Faculté des Sciences, Université Montpellier, Place Eugène Bataillon, 34095 Montpellier, France

ARTICLE INFO

*Article history:*
Received 25 July 2016
Revised 14 September 2016
Accepted 19 September 2016
Available online 20 September 2016

*Keywords:*
Taxonomy
Parvoviruses
Densovirus
Bidensovirus
Circular Rep-encoding ssDNA (CRESS-DNA) viruses
Replication
Recombination and tropism
Viral evolution
Viral chimerism
Review

ABSTRACT

A wide spectrum of invertebrates is susceptible to various single-stranded DNA viruses. Their relative simplicity of replication and dependence on actively dividing cells makes them highly pathogenic for many invertebrates (Hexapoda, Decapoda, etc.). We present their taxonomical classification and describe the evolutionary relationships between various groups of invertebrate-infecting viruses, their high degree of recombination, and their relationship to viruses infecting mammals or other vertebrates. They share characteristics of the viruses within the various families, including structure of the virus particle, genome properties, and gene expression strategy.



**Contents**



* Corresponding authors at: Laboratoire de Virologie (Bldg 18), Institut National de Recherche Scientifique-Institut Armand-Frappier, 531 Boul. des Prairies, Laval, QC, H7V 1B7, Canada (P. Tijssen), Laboratoire de Pathologie Comparée, Faculté des Sciences, Université Montpellier, Place Eugène Bataillon, 34095 Montpellier, France (M. Bergoin).

http://dx.doi.org/10.1016/j.jip.2016.09.005
0022-2011/

## 1. Synopsis of single-stranded DNA virus features

The number of single-stranded DNA (ssDNA) virus families of animals is limited since: (i), ssDNA cannot be transcribed/translated, e.g. to yield DNA polymerase; (ii), linear ssDNA cannot, in contrast to linear single-stranded RNA, be replicated without loss of genomic integrity (corresponding to Okazaki fragments) unless complex hairpin structures, or a protein-primed mechanism (Tijssen and Bergoin, 1995) are introduced to create first a double-stranded DNA (iii), the absence of a complementary DNA strand, as a repair template, increases the mutation rate, and therefore limits the maximum length (2–12 kb with very few exceptions, Mochizuki et al., 2012) to that of RNA viruses (Shackelton et al., 2005). Moreover, these viruses have small icosahedral capsids with a diameter of 18–27 nm with a limited packaging capacity. Although many densoviruses are highly pathogenic, some are essential for the life cycle of the host. Ryabov et al. (2009) demonstrated that infection with *Dysaphis plantaginea* densovirus (DplDV) is essential for the production of the winged morph in asexual clones of the rosy apple aphid and its colony dispersal to neighboring plants. This mutualistic relationship between the rosy apple aphid and this virus results both in a negative impact of DplDV on rosy apple aphid reproduction, but also contribute to the survival of aphid colonies by inducing wing development and promoting dispersal.

Despite the limited number of ssDNA virus families, these viruses are exceptionally widespread, including medically and economically important pathogens (Krupovic, 2013) in all domains of life. Metagenomic studies dramatically boosted our knowledge about ssDNA viruses in the biosphere, from the human gut to hot springs (Rosario et al., 2012). The diversity of ssDNA viruses appears to be determined by two principal factors: extremely high nucleotide substitution rates that approach those of RNA viruses (Shackelton et al., 2005) facilitating adaptation to different environments, extending to the oceans (Labonté and Suttle, 2013), and pervasive recombination, both by DNA and RNA donors (Martin et al., 2011).

Known ssDNA viruses of invertebrates belong to the families *Parvoviridae*, of which the subfamily that infects invertebrates is called *Densovirinae*, *Circoviridae* (usually from the genus *Cyclovirus*) and *Bidnaviridae* (thus far solely in insects). Although different parvovirus genera exist, they have a common Non Structural (*rep*) gene set, coding for proteins that are accessory to Rolling Circle Replication (Section 2.1) at the left half of the genome and a structural capsid protein (*cap*, Section 2.2) gene set at the right hand of the genome. Because host cell DNA polymerase can synthesize the double-stranded molecule as a prelude to transcription/translation only during the S-phase of the cell cycle, these ssDNA viruses have a tropism for rapidly dividing cells which translates in common pathogenic features. The family *Parvoviridae* is a very diverse group of viruses and occurs both in invertebrates and vertebrates, including mammals. The *rep* and *cap* genes of the *Circoviridae* do have a less fixed position with respect to the origin of replication.

## 2. Imperative genes of single-stranded DNA viruses

### 2.1. Rep protein: conserved sequence with initiator proteins for rolling-circle replication (RCR)

An HuHuuu amino acid motif (in which u represents bulky hydrophobic residues) near the N-terminus was identified in the Rep protein (HUH endonuclease superfamily) that is conserved in two vast classes of proteins, one of which is involved in initiation and termination of rolling-circle DNA replication, or RCR (Rep proteins), and the other in mobilization (conjugal transfer) of plasmid DNA (Mob proteins) (Ilyina and Koonin, 1992). An additional conserved motif in this superfamily of proteins is located closer to the N-terminus, and another, downstream of the HuHuu motif, with a conserved Tyr residue. The major role of these HUH endonucleases is processing a range of mobile genetic elements by catalyzing cleavage and rejoining of single-stranded DNA using the active-site Tyr residue in the downstream motif to make a transient 5′-phosphotyrosine bond with the DNA substrate, such as in rolling-circle replication, in various types of transposition and in intron homing (Chandler et al., 2013). HUH enzyme activities require a divalent metal ion, coordination of which is provided by the HUH motif, to facilitate cleavage by locating and polarizing the scissile phosphodiester bond.

Rep proteins also contain a C-terminal superfamily 3 (SF3) 3–5′ helicase domain. RCR uses this ringlike, hexameric 3–5′ helicase activity acting on the template strand to facilitate DNA unwinding at the replication fork (see Section 3.3). The crystal structure of an SF3 DNA helicase, Rep40, from adeno-associated virus 2 (AAV2) has been reported (James et al., 2003) and delineates the expected Walker A and B motifs, but also reveals an unexpected "arginine finger". The so-called arginine finger penetrates the active site of a neighboring subunit. Mutation of the Arg finger to alanine resulted in deficiency in helicase activity. The presence of a functional arginine finger directly implicates the requirement for oligomerization in order to create a competent catalytic center for cooperative ATP hydrolysis (James et al., 2003).

The peptide linker between the HUH endonuclease and the SF3 helicase domains has a critical role in helicase oligomerization (Maggin et al., 2012). The helicase is also required for packaging of the ssDNA genome in the preformed procapsids (King et al., 2001).

### 2.2. Cap protein: homology to that of small, icosahedral RNA viruses

Capsid proteins of small RNA viruses have a so-called β-barrel (or jelly-roll capsid, JRC), consisting of at least 8 β-strands, that are highly conserved. There is strong evidence that these viruses donated the *cap* gene to the T = 1 ssDNA viruses (Koonin et al., 2015a,b). The JRC is indeed highly conserved while the loops between the β-strands are very different from the loops of viruses in other genera. Also, when the icosahedral symmetry axes of two ssDNA viruses are superimposed, the β barrel may have to be rotated and translated radially in order to superimpose them (Simpson et al., 1998). The barrels are located on the inside of the capsid proteins and the loops mostly on the outside of the capsid (Fig. 1).

The JRC consists of 8 β-strands making 2 opposing β-sheets (CHEF and BIDG, naming β-strands from N- to C-terminal βA to βG, respectively). Interestingly, in densoviruses, the βB strand is a linear extension of the N-terminal βA strand, allowing βA to hydrogen bond with βB of the neighboring, twofold-related subunit and thereby increasing the number of interactions between subunits (Simpson et al., 1998). In vertebrate parvoviruses, however, βA folds back and hydrogen bonds with the βB strand from its own subunit ((Fig. 1). The same structural difference was observed for small icosahedral RNA viruses from invertebrates vs those of vertebrates, e.g. human rhinovirus versus cricket paralysis virus (Kaufmann et al., 2010, 2011; Meng et al., 2013). The presence in the same genome of genes with different evolutionary histories (*rep* compared to *cap*) illustrates the pitfalls of genome-based virus classification.

## 3. Densoviruses (subfamily *Densovirinae* of *Parvoviridae*)

### 3.1. Pathology and host range

The family *Parvoviridae* is a very diverse group of viruses and occurs both in invertebrates ("densoviruses") and vertebrates