Virology 514 (2018) 192–202

Contents lists available at ScienceDirect

Virology

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

Functional diversification upon leader protease domain duplication in the *Citrus tristeza virus* genome: Role of RNA sequences and the encoded proteins

Sung-Hwan Kang[1], Osama O. Atallah[1,2], Yong-Duo Sun, Svetlana Y. Folimonova[*]

University of Florida, Plant Pathology Department, Gainesville, FL 32611, USA

ARTICLE INFO

Keywords:
Citrus tristeza virus
Leader protease
Gene duplication
Host range

ABSTRACT

Viruses from the family *Closteroviridae* show an example of intra-genome duplications of more than one gene. In addition to the hallmark coat protein gene duplication, several members possess a tandem duplication of papain-like leader proteases. In this study, we demonstrate that domains encoding the L1 and L2 proteases in the *Citrus tristeza virus* genome underwent a significant functional divergence at the RNA and protein levels. We show that the L1 protease is crucial for viral accumulation and establishment of initial infection, whereas its coding region is vital for virus transport. On the other hand, the second protease is indispensable for virus infection of its natural citrus host, suggesting that L2 has evolved an important adaptive function that mediates virus interaction with the woody host.

1. Introduction

Gene duplication is a major force that generates new genetic material leading to evolutionary innovation. As a result of such events, a duplicated copy, which often is subjected to a lesser selective pressure, can develop a new and different gene function. Alternatively, the gene function could partition among the two copies with certain adaptive benefits. In other situations, when duplication of a gene has no beneficial effect on the host, a second copy may be lost during evolution. While there are many instances of gene duplications in various bacteria and eukaryotes, in viruses such cases have been found less frequently. In particular, RNA viruses, which have severe constraints for their genome sizes, show only a small number of examples. Among those are gene duplication events reported for the reverse-transcribing viruses within the *Retroviridae* family, single-stranded negative RNA viruses from the *Rhabdoviridae* family, and single-stranded positive RNA viruses from a few families (Forss and Schaller, 1982; Tristem et al., 1990; Walker et al., 1992; Kambol et al., 2003; Valli et al., 2006, 2007; Simon-Loriere and Holmes, 2013; Willemsen et al., 2016). Strikingly, viruses from the family *Closteroviridae*, which combines plant viruses with the largest RNA genomes ranging from ~ 15.5 to ~ 19.3 kb and encoding 10–14 proteins, exemplify the case of intra-genome duplications of more than one gene. All viruses belonging to this family carry a tandem duplication of the coat protein gene as a part of their hallmark quintuple gene block, which is unique to *Closteroviridae* (Boyko et al., 1992; Karasev et al., 1995; Dolja et al., 2006). Additionally, a number of members of the genus *Closterovirus* possess a tandem of papain-like cysteine proteases that are also thought to have evolved via gene duplication (Karasev et al., 1995; Karasev, 2000; Peng et al., 2001; Dolja et al., 2006; Liu et al., 2009). These two proteases are encoded in the 5′-terminal region within the open reading frame 1a (ORF 1a) upstream of the conserved replicase domains and represent a class of viral papain-like "leader" proteases, which are found among diverse positive-strand RNA viruses and proved to be involved in different aspects of virus-host interactions. Besides autocatalytic processing, leader proteases of plant and animal viruses were shown to function in virus replication, virion assembly, cell-to-cell and systemic movement, vector transmission, viral superinfection exclusion, and circumvention of a host RNA silencing-based defense response (Gorbalenya et al., 1991; Kasschau et al., 1997; Kasschau and Carrington, 1998; Ziebuhr et al., 2000; Peng and Dolja, 2000; Peng et al., 2001, 2003; Santos et al., 2006; Ng and Falk, 2006; Liu et al., 2009; Atallah et al., 2016; Valli et al., 2017). Interestingly, duplications of leader proteases have been also reported for some animal viruses with largest known RNA genomes (from ~ 26 to ~ 32 kb) united in the order *Nidovirales*. Some of these viruses possess either duplicated or triplicated adjacent papain-like protease domains (Ziebuhr et al., 2000; Gorbalenya et al., 2006).

We have been working with *Citrus tristeza virus* (CTV), one of the most complex members of the family *Closteroviridae* (Bar-Joseph et al., 1979; Dolja et al., 1994, 2006; Agranovsky, 1996, 2016; Karasev,

[*] Corresponding author.
E-mail address: svetlana@ufl.edu (S.Y. Folimonova).
[1] S.-H. Kang and O.O. Atallah contributed equally to this work.
[2] Current address: Zagazig University, Department of Plant Pathology, Zagazig 44519, Egypt.

https://doi.org/10.1016/j.virol.2017.11.014
Received 24 October 2017; Received in revised form 15 November 2017; Accepted 17 November 2017

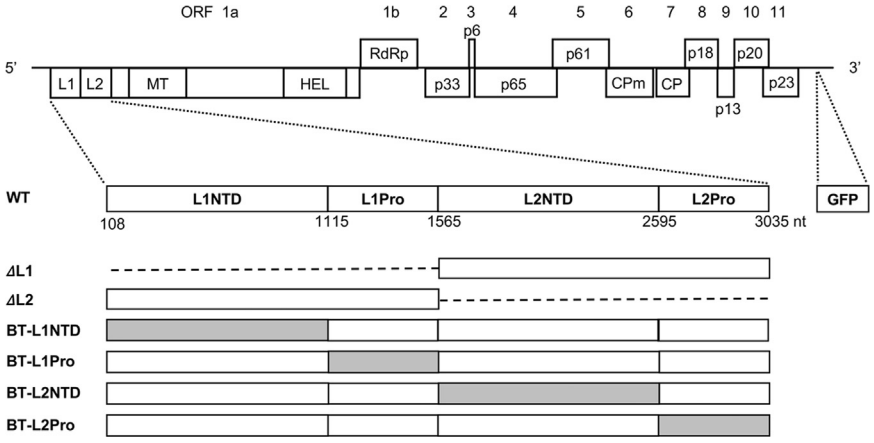

Fig. 1. Schematic representation of the genome of CTV and its variants generated in this study. The boxes represent ORFs and their translated products. L1 and L2, papain-like leader proteases; NTD, N-terminal domain; Pro, C-terminal proteolytic domain; MT, methyltransferase; HEL, helicase; RdRp, RNA-dependent RNA polymerase; p65, HSP70 homolog; CPm, minor coat protein; CP, coat protein. Schemes of virus variants created by deletion (shown as dashed lines) or substitution (shown as grey boxes) of genomic fragments within the L1 or L2 coding regions are shown under the enlarged genome segment. The nucleotide positions for each domain within L1 and L2 are indicated. All viruses have the GFP ORF inserted between the p23 gene and the 3′ non-translated region of the CTV genome under the native promoter of the CTV CP subgenomic RNA.

2000). CTV has a 19.3 kb positive-sense RNA genome organized into twelve ORFs, which encode proteins functioning at different stages of the virus cycle (Fig. 1; Pappu et al., 1994; Karasev et al., 1995; Karasev, 2000). The first ORF expressed from the genomic RNA encodes a polyprotein, which carries two papain-like leader protease domains L1 and L2 positioned at its N-terminal end along with methyltransferase- and helicase-like domains. Occasional translation through the ORF1b mediated by a +1 frameshift results in the production of a larger polyprotein, which contains an RNA-dependent RNA polymerase-like domain (Karasev et al., 1995). Ten ORFs located in the 3′ portion of the virus genome are expressed via 3′-coterminal subgenomic RNAs (Hilf et al., 1995) and encode major (CP) and minor (CPm) coat proteins, p65 (a homolog of cellular HSP70 proteins), and p61 that are required for virion assembly (Satyanarayana et al., 2000) and movement along with the hydrophobic p6 protein (Dolja et al., 2006; Tatineni et al., 2008); p20 and p23 proteins that function as suppressors of the host RNA silencing along with CP (Lu et al., 2004); and three proteins p33, p18, and p13, which are thought to play a role in extending the virus host range (Tatineni et al., 2008, 2011).

In addition to CTV, several other viruses representing the genus *Closterovirus* have been studied (Zhu et al., 1998; Dolja, 2003; Tzanetakis et al., 2005, 2007; Tzanetakis and Martin, 2007). Among those, the role of viral leader proteases in the virus infection cycle has been examined for *Beet yellows virus* (BYV) and *Grape leaf-roll associated virus 2* (GLRaV-2) (Peng and Dolja, 2000; Peng et al., 2001, 2003; Liu et al., 2009). Similarly to CTV, GLRaV-2 encodes two leader proteases (L1 and L2), while BYV possesses only one (L-Pro). Each of these proteases contains a variable N-terminal domain and a conserved C-terminal proteolytic domain that mediates the autocatalytic cleavage of the protease from the polyprotein (Peng et al., 2001). For BYV, it was demonstrated that release of L-Pro driven by the C-terminal domain is essential for virus viability, and, moreover, both the N- and C-terminal domains are needed for efficient RNA replication as well as for systemic movement of BYV (Peremyslov et al., 1998; Peng and Dolja, 2000; Peng et al., 2003). Interestingly, in the experiments with chimeric BYV constructs only L1 of CTV, but not L2, was able to substitute for the BYV L-Pro and rescue virus accumulation, though, with a significantly lower efficiency, thus, providing an indication of potential functional divergence of the L1 and L2 domains (Peng et al., 2001). Further analysis of GLRaV-2 deletion mutants showed that although the L1 and L2 domains of GLRaV-2 have partially overlapping functions and demonstrate certain functional redundancy, the former has a more prominent role in virus replication and is essential for establishment of virus infection, while the latter domain plays an accessory part (Liu et al., 2009). The involvement of both domains in virus establishment in the initially inoculated cells was more pronounced upon agroinfiltration of mini-replicon virus variants into the natural grapevine host, compared to that in the laboratory host *Nicotiana benthamiana* (Liu et al., 2009). This approach, however, limited examination of the GLRaV-2 L1 and L2 effects to infiltrated epidermal and mesophyll cells in *Vitis vinifera* and did not allow evaluation of their roles in the development of the systemic infection. Nevertheless, the observed differences led to a hypothesis that the duplicated domains have evolved to expand the host range of the virus (Liu et al., 2009). It remained unclear, though, whether those processes are mediated entirely by the two proteins or involve their coding regions at the RNA level as well.

In this work, we extended the investigation of functional specialization of the tandem papain-like proteases of closteroviruses by examining the involvement of the L1 and L2 proteases encoded in the genome of CTV and their corresponding RNA sequences in virus accumulation and ability to establish systemic infection in the herbaceous host *N. benthamiana* as well as in the natural *Citrus macrophylla* host. By utilizing full-length virus mutants in which the L1 and L2 coding sequences were deleted or altered in a way that production of the wild-type proteins was preserved, we demonstrate a marked functional divergence of the L1 and L2 domains of CTV, which appears to be more striking, compared to the functional specialization of the corresponding domains of GLRaV-2. We show that the L1 domain, specifically, the encoded protease but not the coding RNA plays a crucial role in the establishment of virus infection and virus accumulation. At the same time, the nucleotide sequence of this domain appears to be indirectly involved in virus movement by encompassing the elements needed for the formation of proper virions. On the other hand, the second protease essentially lost the ability to support virus multiplication. Furthermore, rather than being accessory as in case of GLRaV-2, the L2 protease of CTV carries out an indispensable role in mediating virus systemic infection in the natural citrus host.

2. Results

2.1. L1 region mediates the establishment of virus infection and viral RNA accumulation at the protein level, while its RNA sequence is dispensable

To examine the roles of CTV proteases and their coding regions in virus ability to infect plant hosts, we generated six full-length CTV variants with alterations within the corresponding areas of the virus genome (Fig. 1). All engineered viruses contained an extra ORF of the green fluorescent protein (GFP) to facilitate observation of virus accumulation in the inoculated plants. Those included two deletion mutants lacking the entire L1 or L2 coding sequences (ΔL1 and ΔL2, respectively) and four mutants carrying substitutions in the wobble position of each possible codon within the sequences encoding a non-conserved N-terminal domain or a proteolytic C-terminal domain of either protease (BT-L1NTD, BT-L1Pro, BT-L2NTD, and BT-L2Pro; Fig. 1; Supplementary Fig. S1; see Methods). Modifications created in the latter set of virus mutants were distributed throughout the respective domains and were