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

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

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

*Corresponding author. *E-mail address:* svetlana@ufl.edu (S. Y. Folimonova)

[1] Current address: Zagazig University, Department of Plant Pathology, Zagazig 44519, Egypt

S.-H. Kang and O. O. Atallah contributed equally to this work.

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.

Keywords: Citrus tristeza virus, Leader protease, Gene duplication, Host range

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; Peng et al., 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; Agranovsky, 1996; Karasev, 2000; Dolja et al., 2006; Agranovsky, 2016). 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 homologue 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 minireplicon 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.

Results

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 expected to alter primary and secondary structure of the RNA sequences within the L1 and L2 coding regions yet preserve the original amino acid composition of the produced proteins. Alignment of the nucleotide sequences of the BT-L1NTD, BT-L1Pro, BT-L2NTD, and BT-L2Pro viruses showed 74.5, 77.2, 72.8, and 75.4% nucleotide identity with the sequences of corresponding domains present in the wild type (WT) virus (Supplementary Fig. S1). Further analysis showed that these mutations were sufficient to significantly alter the predicted secondary structures of the respective genomic regions in the engineered virus variants as well (Supplementary Fig. S2). Retention of the introduced substitutions in the genomes of the generated mutants during *in planta* propagation (these experiments are discussed below) was verified by sequence analysis of their progeny produced upon inoculation into *N*. *benthamiana* and *C*. *macrophylla* plants as described in the Materials and Methods. Importantly, all sequenced regions in the assayed virus progeny samples were identical to those in the respective original mutant viruses, confirming that no reversions to the wild type have occurred.

As the next step, CTV variants were introduced into *N*. *benthamiana* plants by agroinfiltration using *Agrobacterium tumefaciens* cultures transformed with binary vectors each carrying a cDNA of a mutant virus variant under the *Cauliflower mosaic virus* 35S promoter. Observation of the inoculated leaves at six days post infiltration (dpi) with a hand-held UV light revealed that all virus variants except ΔL1 mutant exhibited strong GFP fluorescence at the levels similar to that of the WT virus indicating robust accumulation of most viruses (Fig. 2A). Further examination of the infiltrated leaves by fluorescence microscopy showed presence of a

few scattered fluorescing cells in ΔL1-infiltrated leaves, which exhibited much lower levels of GFP, compared to cells replicating the parental WT virus, and those were apparently below the level of the hand-held UV lamp detection (Fig. 2B). This indicated that deletion of the L1 coding region resulted in a dramatic decrease in virus ability to accumulate in the primary inoculated cells and establish initial infection. To quantify replication efficiency of individual virus variants, we performed real-time quantitative reverse transcription-PCR (RT-qPCR) analysis using total RNA extracts from the infiltrated leaves. Absolute quantification of viral RNA accumulation showed that prevention of the L1 protease production resulted in a tremendous drop in virus titer to 0.28% compared to that of the WT virus (Fig. 2C). At the same time, mutants in which the coding RNA sequence of the L1 region was altered without affecting the protein sequence (BT-L1NTD and BT-L1Pro) accumulated at the levels that showed no statistically significant difference with the WT virus accumulation (Fig. 2C; Tukey's honest significant difference (HSD) test *post hoc*, $P < 0.01$). On the other hand, lack of L2 did not affect accumulation of CTV at all: similar virus loads were found in leaves infiltrated with ΔL2 as well as with BT-L2NTD and BT-L2Pro constructs (Fig. 2C). These results suggested that expression of the L1 protease but not L2 is essential for virus ability to establish initial infection. Furthermore, these processes appeared to be mediated solely at the protein but not at the RNA level. Remarkably, nucleotide substitutions introduced throughout a large area of virus genome encompassing ~3000 nts, which greatly changed its primary and secondary structure, were well tolerated by the virus replication machinery, provided they did not prevent production of L1.

RNA sequence of the L1NTD domain is involved in virus transport. Experiments described above demonstrated that CTV accumulation in the inoculated cells requires production of the L1

protease, yet, is tolerant to "silent" changes in the L1 coding nucleotide sequence. To assess the effect of those changes in the RNA sequence on virus ability to move in plant hosts, *N. benthamiana* plants infiltrated with the WT-, ΔL1-, BT-L1NTD- or BT-L1Pro-expressing constructs were monitored for spread of GFP fluorescence to upper leaves over a two-month period following inoculation (Fig. 3A). Around three weeks post infiltration (wpi), the appearance of GFP fluorescence in a form of scattered dots along the main veins of upper leaves, which were not infiltrated previously, was noted in the plants inoculated with WT and BT-L1Pro (data not shown). By six wpi, these plants displayed strong GFP fluorescence in veins of different classes in most upper leaves, indicating that the WT and BT-L1Pro viruses have moved systemically (Fig. 3A). As we expected based on inefficient accumulation of the virus variant carrying the deletion of the whole L1 region, no spread of GFP fluorescence was detected in plants inoculated with this virus. At the same time, GFP fluorescence was not observed in the systemic leaves of plants infiltrated with BT-L1NTD either, despite the fact that this virus was accumulating in the primary inoculated leaves at levels similar to that of the WT virus (Figs. 2A, 2C and 3A). Unlike BT-L1Pro, which had alterations of the nucleotide sequence of the coding region of the C-terminal proteolytic domain of L1 yet successfully invaded the inoculated plants, BT-L1NTD was unable to move long distance. To confirm systemic spread of the WT and BT-L1Pro viruses and lack of that upon inoculation with ΔL1 or BT-L1NTD, we analyzed extracts obtained from systemic leaves of the inoculated plants by conventional RT-PCR and ELISA. Total RNA extracts from systemic leaves of plants infiltrated with WT and BT-L1Pro subjected to RT-PCR using primers specific to the CTV CP coding region showed amplification of the corresponding fragment of ~800 nts in size (Fig. 3B). No product amplification occurred from RNA extracted from systemic leaves collected from the ΔL1- or BT-L1NTD-infiltrated plants,

confirming lack of virus accumulation (Fig. 3B). Similar results were obtained using a serology-based analysis with the CTV virion-specific antiserum (Fig. 3C). Virus accumulation in the systemic tissue in the BT-L1Pro-infected plants was similar to that in the WT-infected plants. On the other hand, ELISA values obtained for systemic leaves collected from plants infiltrated with ΔL1 or BT-L1NTD were as low as that of healthy plants (Tukey's HSD test *post hoc*, $P < 0.01$).

As the next step, we tested the ability of L1 mutants to infect citrus, the natural host of CTV. Following our routine approach for introducing recombinant CTV variants into citrus seedlings, virion preparations made using infiltrated *N*. *benthamiana* leaves were used for bark-flap inoculation of *C*. *macrophylla*, one of the most susceptible citrus varieties. The ΔL1 virus was not included in this assay due to its greatly debilitated accumulation in the infiltrated leaves and lack of ability to systemically infect *N*. *benthamiana* plants. In addition, as discussed below, no virions were found in preparations obtained from leaves infiltrated with this mutant. At three months after inoculation, newly emerged tissue was collected and examined for virus accumulation. Strong GFP fluorescence was detected in the phloem cells of citrus trees inoculated with WT and BT-L1Pro, indicating successful systemic movement of both viruses (Fig. 3D). No GFP fluorescence was observed in trees inoculated with BT-L1NTD. RT-PCR and ELISA analyses conducted as described above confirmed the establishment of systemic infections and lack of those in the plants inoculated with the respective viruses (Fig. 3D and E). Similarly to what was observed in *N*. *benthamiana*, a laboratory herbaceous host, the virus mutant carrying alterations within the coding sequence of the N-terminal domain of L1, which significantly altered its primary and secondary structure yet preserved the protein sequence, was unable to infect the natural virus host. On the other hand, similar changes in the coding region of the proteolytic C-terminal domain did not affect virus infectivity. These results suggested that the

RNA sequence in the coding region of the non-conserved N-terminal domain of L1 is essential for virus movement ability.

L2 region mediates virus systemic infection at the protein level in a host-dependent manner. Our next goal was to examine a role of CTV L2 protease and its coding RNA sequence in virus ability to systemically infect both types of hosts. *N*. *benthamiana* plants infiltrated with CTV variants carrying mutations in the L2 region (ΔL2, BT-L2NTD, and BT-L2Pro) were monitored for the appearance of GFP fluorescence in the developing systemic tissues over a two-month period. Interestingly, the phenotypes of the L2 mutants were quite distinct from those produced by the mutants having alterations within the L1 domains. At three wpi, GFP fluorescence was observed in plants infected by the WT, BT-L2Pro, and, importantly, BT-L2NTD viruses in a form of scattered dots along the plant stem and main veins of the upper leaves (data not shown). Three weeks later, these plants were showing GFP fluorescence along the veins in most upper leaves, indicating the efficient long distance spread of viruses with multiple silent mutations in the L2 coding region (Fig. 4A). Remarkably, unlike plants infiltrated with ΔL1 that have not become infected with the mutant, plants inoculated with ΔL2 displayed GFP fluorescence in the upper leaves, though, it appeared with some apparent delay of about two weeks compared to the emergence of GFP fluorescence in plants inoculated with the WT, BT-L2NTD, and BT-L2Pro viruses (Fig. 4A). In addition to showing slower long distance movement, ΔL2 appeared to have decreased invasiveness: the number of fluorescing sites in the ΔL2-infected systemic leaves was greatly reduced (Fig. 4A and B). Further RT-PCR and ELISA tests demonstrated that BT-L2NTD and BT-L2Pro accumulated at the levels similar to that of the WT virus (Fig. 4C and D; Tukey's HSD test *post hoc*, P values < 0.01). At the same time, the

accumulation of ΔL2 in the systemic leaves was greatly diminished, which corroborated with the observations of the expression of GFP fluorescence discussed above (Fig. 4C and D). Additionally, there was an obvious attenuation of the symptoms produced by the ΔL2 virus compared to those produced by other mutant viruses examined here. No typical vein-clearing that is usually seen in the systemic leaves of CTV-infected *N*. *benthamiana* plants was noted upon ΔL2 infection (Fig. 4E). Furthermore, plants infected with ΔL2 showed much healthier appearance compared to plants infected with WT or other mutants (Fig. 4F).

Next, we assessed the effect of alterations within the coding region of L2 on virus infection of citrus plants. Following the procedure described above, we introduced mutant viruses into young seedlings of *C*. *macrophylla* using virions extracted from infiltrated *N*. *benthamiana* leaves (note that those infiltrated leaves of *N*. *benthamiana* used as inoculum sources showed similar viral loads for all virus variants tested in this experiment as shown in Fig. 2A and C). Remarkably, silent mutations introduced throughout the whole L2 coding region did not affect virus ability to infect the citrus host either: both BT-L2NTD and BT-L2Pro established systemic infections in the inoculated citrus plants within the same time intervals as the WT virus (Fig. 4G). Furthermore, both viruses accumulated at the levels comparable to that of the WT (Fig. 4H and I; Tukey's HSD test *post hoc*, P values < 0.01). On the other hand, ΔL2, which was able to spread systemically in the herbaceous host, showed complete inability to develop infection in citrus. No GFP fluorescence was detected in tissue samples from the ΔL2-inoculated *C*. *macrophylla* plants (Fig. 4G). Microscopy observations were supported by following RT-PCR and ELISA tests showing lack of virus accumulation in the latter plants (Fig. 4H and I). Collectively, these results showed that the L2 region is involved in systemic infection of host plants. Furthermore, such involvement is host-dependent and is mediated by the L2 protein and

not by its coding RNA sequence. While the L2 protease plays a supplementary role in the herbaceous *N. benthamiana* host, it is indispensable for the infection of citrus.

Movement deficiency of a virus carrying alterations in the coding sequence of the L1NTD domain correlates with improper virion assembly. Closteroviruses form long flexuous virions with a complex bipolar structure (Agranovsky et al., 1995; Febres et al., 1996; Tian et al., 1999). CTV virions are about 2000 X 12 nm in size and are built of two coat proteins: the CPm encapsidates the 5' 630 nts of the genomic RNA, and the CP encapsidates the remainder portion of the genome (Satyanarayana et al., 2004). Furthermore, it was shown that interaction of CPm and its assembly region displays certain degree of specificity (Tatineni et al., 2010). In addition, two other viral proteins – p65 and p61 – participate in assembly of virions by restricting CPm encapsidation to the 5'-terminal region, and, likely, they represent minor integral components of the virion tails (Napuli et al., 2000; Alzhanova et al., 2001, 2007; Satyanarayana et al., 2004). It has been hypothesized that p61 and p65 may have a recognition signal within this 5'-genomic region and form a "lock" at the interface between the CPm- and CP-coated portions (Agranovsky, 2016). In relation to our study, it is important to emphasize that this transition zone lies within the coding sequence for the N-terminal domain of L1. Thus, alteration of this region could negatively affect formation of proper virions, which would have a detrimental effect on virus movement, since closteroviruses are known to move in a virion form (Cronshaw et al., 1966; Peng et al., 2003; Dolja et al., 2006; Alzhanova et al., 2007). Therefore, we next examined virion formation ability of the mutant viruses generated in this work. Virus particles were extracted from infiltrated *N. benthamiana* leaves and examined using transmission electron microscopy (TEM; Fig. 5A). Virions isolated from the WT-infected plants were clearly seen by

negative staining (Fig. 5A; WT). The average length of the virions observed from the WT sample was 2,476 nm (Fig. 5B; note that the genomes of the viruses used in this work contained extra ORF, that of GFP, positioned under the added CTV CP subgenomic RNA promoter, which results in formation of longer virions, compared to those formed by the native non-tagged CTV). Among mutants, three viruses, BT-L1Pro, BT-L2NTD, and BT-L2Pro, which successfully established systemic infections in both plant hosts, showed virions with the average lengths being comparable to the average length of the WT virions (Fig. 5A and B). Furthermore, no virions were found in preparations obtained from ΔL1-infiltrated leaves (data not shown). Virions produced by the ΔL2 virus were seen as distinctive CTV-like particles. However, as we expected, the size of those virions was reduced compared to virions formed by WT (Fig. 5A and B), which apparently resulted from shortening the genomic RNA due to the deletion of the L2 coding region. Formation of virus particles by this mutant variant correlated with its movement capability observed in *N*. *benthamiana* plants (Fig. 4A). On the other hand, BT-L1NTD, which genome size was not affected by introduced alterations, also formed shorter-than-full-length particles. The average length of those was 2,173 nm that was significantly less than the average lengths of WT and the other three mutant viruses, which genomes had exactly the same length (Fig. 5B; Tukey's HSD test *post hoc*, P values < 0.01). Therefore, BT-L1NTD appeared to have certain deficiency in virion formation, which resulted from the mutations introduced in the region shown to be involved in virion assembly (discussed above). It is likely that inability to form complete virions prevented this virus from systemic transport in the host plants.

L1 and L2 proteases show differential subcellular localization. Subcellular localization of a protein often relates to its function. The results of our work presented above demonstrated

functional divergence of the tandem leader proteases of CTV. Therefore, we examined their intracellular distribution in order to determine whether certain differences between L1 and L2 could be found at this level as well. *A. tumefaciens* cultures transformed with binary vectors carrying a fusion of the cyan fluorescent protein (CFP) gene to the L1 ORF (CFP:L1) or a fusion of the red fluorescent protein (RFP) gene to the L2 ORF (RFP:L2) positioned under the 35S promoter were infiltrated into *N. benthamiana* leaves and examined using confocal laser scanning microscopy at four dpi. Both FP-tagged proteases produced bright fluorescence allowing the evaluation of their distribution inside living cells. Most fluorescence of CFP:L1 was found in multiple punctate spots along the cellular peripheral membrane. Those spots appeared to be stationary as they remained motionless during observation (Fig. 6A). At the same time, RFP:L2 was usually seen as a single cytoplasmic granule that had a shape of an elongated sphere (Fig. 6A). Apparently, the RFP:L2 bodies were mobile as they were constantly moving inside the cells (see time-lapse micrographs in Fig. 6B; Supplementary video S3). Analysis of the aa sequences of the L1 and L2 proteins revealed presence of a hydrophobic region overlapping with a transmembrane domain in the middle portion of L1, which corroborated with the detected membrane localization of this protease. In contrast, no transmembrane domain was predicted for L2 (Fig. 6C). The observations obtained here suggested that the L1 and L2 proteases of CTV have distinct intracellular distribution and appear to be targeted to different cellular compartments, supporting the idea of the diversification of their functions in virus infection.

Discussion

Earlier studies suggested that independent gene duplication events are responsible for the emergence of the tandem domains coding for the papain-like leader proteases found in a number of closteroviruses (Peng et al., 2001). It was noted that the emerged paralogs could have undergone some divergence in their sequences and functions (Peng et al., 2001; Liu et al., 2009). It was not clear, however, what is the degree of such divergence. For instance, while certain specialization of the two leader protease domains of GLRaV-2 was demonstrated, it was also shown that their roles in the viral cycle have redundancy (Liu et al., 2009). It has remained to be determined whether the specific processes are mediated by the encoded proteins or their RNA sequences. Here we showed that the L1 and L2 proteases encoded in the CTV genome play important yet highly distinctive roles in the virus infection cycle, which correlates with limited conservation of their aa sequences. We also addressed the question of the involvement of the coding nucleotide sequences of the respective proteins in the virus processes.

The observations made in this work demonstrated that expression of the L1 protease is crucial for virus ability to accumulate in the initially infected cells. On the other hand, the second protease of CTV, L2, had practically lost its ability to support virus accumulation. Upon deletion of the L1 sequence, virus accumulation was recovered at only 0.28% compared to that of the WT virus. At the same time, deletion of L2 had no effect on virus accumulation. This differs from what was reported previously for the leader protease domains of another closterovirus, GLRaV-2 (Liu et al., 2009). In that case, L2 of GLRaV-2 alone was able to support virus accumulation at 22%. Thus, our study indicates that in comparison with the L1 and L2 domains of GLRaV-2 the respective domains of CTV have undergone more extensive functional diversification.

Furthermore, it deciphers the role of the encoded proteins versus their RNA sequences. As we demonstrated, the involvement of the L1 domain in virus accumulation appears to be solely mediated at the protein level. Silent alterations throughout the coding sequence of L1 did not result in the change of viral accumulation. On the other hand, such modifications highlighted a role of the RNA sequence corresponding to the N-terminal sub-domain of L1 in the assembly of viral particles and, therefore, in virus long-distance transport. The inability of the respective mutant virus (BT-L1NTD) to move systemically in the inoculated plants correlated with its deficiency in virion formation, which corroborated with the previous work showing that this region of the genome encompasses structural elements involved in assembly of proper virion tails of CTV (Satyanarayana et al., 2004). Therefore, it appears that introduced mutations affected virus movement indirectly by affecting formation of virions.

Experiments that involved virus mutants carrying deletions or alternations within the coding region of L2 demonstrated that production of this protease 1) improves virus fitness in the herbaceous host and 2) is crucial for virus ability to infect *Citrus* spp. Here, using full-length viruses capable of systemically infecting plant hosts and an established procedure for inoculation of citrus trees we were able to overcome limitations of an earlier study with GLRaV-2, which examined ectopic amplification of truncated minireplicon variants in agroinfiltrated epidermal and mesophyll cells of grapevine leaves without assessing the role of GLRaV-2 proteases in the systemic infection of the viral natural host (Liu et al., 2009). Thus, our work extended previous observations and, moreover, brought about the discovery that L2 of CTV is indispensable for virus systemic infection in the natural host. In our experiments, the ΔL2 mutant completely failed to infect *C. macrophylla*, one of the most CTV-susceptible citrus varieties. Furthermore, silent changes in the corresponding RNA sequence did not show any effect on virus biology,

indicating that the underlying processes are regulated at the L2 protein level. One possible mechanism is that L2 operates by counteracting plant defense responses. Being less critical for infection of *N. benthamiana*, a plant species known to be highly permissive for many different viruses, this function of L2 becomes vital for virus infection of the natural woody host. Thus, it appears that the role of L2 is to assist the infection of plant hosts by mediating virus-host interactions. It is worth mentioning that, in addition to CTV and GLRaV-2, the leader protease duplication is also found in other closteroviruses like *Strawberry chlorotic fleck virus* (Tzanetakis and Martin, 2007), *Raspberry leaf mottle virus* (Tzanetakis et al., 2007), and *Rose rosette-associated virus* (He et al., 2015), which infect perennial hosts as well, but not in BYV that infects herbaceous annual hosts. Somewhat similar situation has been found with ipomoviruses of the family *Potyviridae* in which the P1 leader protease duplication has facilitated adaptation of these viruses to a wide range of plant hosts (Valli et al., 2007).

Functional differences in the L1 and L2 proteases of CTV appear to be in agreement with differences found in their intracellular localization and biochemical properties such as presence of a transmembrane domain in the N-terminal portion of L1 driving its association with the cellular membrane and lack of that in the second protease. These observations are in the agreement with the previous comparative analysis of the L1 and L2 aa sequences, which demonstrated that sequence conservation is limited to their C-terminal catalytic domains encompassing residues that participate in proteolysis, whereas N-terminal domains share no common aa motifs (Karasev et al., 1995; Peng et al., 2001).

Our work illustrates marked functional diversification of the tandem domains of the leader proteases of CTV, which have evolved via gene duplication event, and addresses the roles of the coding RNA sequences and the corresponding proteins in the virus infection cycle.

Furthermore, our results along with the fact that leader protease duplications are found in closteroviruses infecting perennial hosts, but not in those that only infect herbaceous plant species allow us to hypothesize that the acquisition of the second protease domain and its further diversification allowed CTV to adapt to a new niche represented by a woody host with a long life span. We trust that observations presented in this work further our understanding of the complex biology and evolution of viruses with large RNA genomes.

Materials and Methods

Generation of virus constructs. The constructs of all CTV mutants used in this study (pΔL1, pΔL2, pBT-L1NTD, pBT-L1Pro, pBT-L2NTD, and pBT-L2Pro) were generated using an infectious cDNA clone of the T36 isolate of CTV tagged with the GFP ORF placed under the cauliflower mosaic virus 35S promoter in the binary vector pCAMBIA1380 (pWT herein; El-Mohtar and Dawson, 2014). For pBT-L1NTD, pBT-L1Pro, pBT-L2NTD, and pBT-L2Pro, the coding sequence within the target domains (spanning nts 108-1115, 1116-1565, 1566-2594, and 2595-3035, respectively) was mutated by substituting nts in the wobble position of each possible triplet (Fig. 1 and Supplementary Fig. S1). In order to do this, L1 and L2 proteins were back-translated using EMBOSS_Backtranseq tool (http://www.ebi.ac.uk/Tools/st/emboss_backtranseq/) and subjected to codon optimization. The new nucleotide sequences of the L1 and L2 domains plus the flanking 5' and 3' viral sequences spanning between the *Asc*I and *Bsu36*I restriction sites within the CTV T36 infectious clone were synthesized through GenScript custom gene synthesis service and cloned into the pUC58 vector (GenScript). The synthesized fragments in the pUC58 vector were digested using *Asc*I and

*Bsu*36I restriction enzymes and ligated into the corresponding region of pWT, which was digested with the same restriction enzymes. Clones were screened by restriction enzyme digestion pattern and confirmed by sequencing. The pΔL1 and pΔL2 deletion mutants were generated by overlapping extension PCR (OE-PCR). For pΔL1, two flanking PCR fragments generated using a set of *Asc*I FW (5′-AAAGGCGCGCCGAATTCGATCTCCTTTGCCCC-3′) and ΔL1 RV (5′-GCCCAATTAAGACGTGGGCAGACGTCATAGTGTAATTTTTGTTGTGGG-3′) primers for the upstream fragment and a set of ΔL1 FW (5′ CCCACAACAAAAATTACACTATGACGTCTGCCCACGTCTTAATTGGGC-3') and *Bsu*36I RV (5'- CCACTATTCCACCGAAACCCGGT-3') primers for the downstream fragment were used as templates for OE-PCR performed using a set of *Asc*I FW and *Bsu*36I RV primers. The obtained product was digested with *Asc*I and *Bsu*36I restriction nucleases and substituted for the corresponding fragment in the pWT infectious clone. The same strategy was used for the generation of pΔL2 that was carried out using ΔL2 FW (5′-CCATCATTTAGTTGGTGGGGAAGATAACTCCCTGATAATCACAG-3') and ΔL2 RV (5'-CTGTGATTATCAGGGAGTTATCTTCCCCACCAACTAAATGATGG-3') primers for PCR. Clones were screened by restriction enzyme digestion pattern and confirmed by sequencing. Engineered CTV variants were used for inoculation of *N*. *benthamiana* and *C*. *macrophylla* plants as described below. Maintenance of the introduced substitutions in the progeny of the generated virus variants has been verified by sequence analysis of the RT-PCR products obtained

using total RNA extracts from virus-inoculated plants and primers corresponding to the modified regions (primers are available upon request). Sequence analysis was performed at Macrogen (Maryland, USA).

Generation of fluorescent protein-tagged leader protease constructs. The pCFP:L1 and pRFP:L2 constructs were created in pCASS-4N by OE-PCR using two PCR fragments, one containing a fusion of the 5' non-translated region of tobacco etch virus (TEV; Carrasco et al., 2007) to the CFP (or RFP) ORF and the other containing the complete ORF of L1 (or L2) similar to a strategy described previously[57]. The first fragment, TEV-CFP (or TEV-RFP), was generated by fusing the 5' non-translated region of TEV in front of the CFP (or RFP) ORF using PCR with a set of a forward primer (5'- AACACAACATATACAAAACAAACGAAT-3') and a reverse primer (5'- GCTTCCTCTGAGTTTCGACATCTTGTACAGCTCGTCCATGCCG-3' for the CFP ORF or 5'- AATTAAGACGTGGGCAGACGTCTTGTACAGCTCGTCCATGCC-3' for the RFP ORF). The second fragment encompassing the L1 (or L2) coding sequence was generated by PCR with a set of a forward primer (5'-ATGGACGAGCTGTACAAGATGTCGAAACTCAGAGGAAGC-3' for L1 or 5'-AACACAACATATACAAAACAAACGAAT-3' for L2 and a reverse primer (5'-TTACCATGGTCACCCACCAACTAAATGATGGTTAGG-3' for L1 or 5'-TTACCATGGTCAACCACCCATCTTATGGTACCCATTA-3' for L2). The OE-PCR was performed using C-1728 (5'- AACACAACATATACAAAACAAACGAAT-3') and C-2171 (5' TTACCATGGTCACCCACCAACTAAATGATGGTTAGG-3') to create pCFP:L1. The OE-PCR was performed using C-1728 and C-2174 (5'-TTACCATGGTCAACCACCCATCTTATGGTACCCATTA-3') to create pRFP:L2. Vent DNA

polymerase (New England Biolabs, Inc.) was used for the OE-PCR in order to generate blunt-ended amplicons. The PCR products were digested with *Nco*I restriction endonuclease. The digested products were substituted for the corresponding fragment in the pCASS-4N plasmid, which was also digested with *Stu*I and *Nco*I restriction endonucleases.

Agroinfiltration of CTV constructs into *N. benthamiana*. Agroinfiltration of constructs was conducted as previously described (Carrasco et al., 2007; Ambrós et al., 2011; Kang et al., 2015). Briefly, plasmids were introduced by heat shock into *A. tumefaciens* EHA105, and the resulting transformants were selected on the Luria-Bertani agar plates containing two antibiotics (50 μg/ml rifampicin and 25 μg/ml kanamycin). Cells collected from an over-night culture of a selected single colony were gently resuspended in a buffer containing 10 mM 2-(N-morpholino) ethane sulfonic acid (MES, pH 5.85), 10 mM $MgCl_2$ and 150 mM acetosyringone at $O.D._{600nm}$ = 0.1 or 1.0 (for clones constructed in pCAMBIA1380 or pCASS-4N plasmid, respectively). Following three-hour incubation at room temperature without shaking the suspension was infiltrated into three-week-old *N. benthamiana* plants using needless syringe.

Inoculation of citrus plants. The infiltrated leaves of *N. benthamiana* plants that tested positive for CTV infection by RT-PCR and fluorescence microscopy were harvested at seven dpi and used for virion extraction as described by Robertson et al., 2005. The virion preparations then were used for "bark-flap" inoculation of nine to 12 months old trees of *C. macrophylla*. The inoculated citrus plants were maintained under the greenhouse conditions until further analysis.

Total RNA extraction. 150 mg of bark tissue collected from young flushes of *C*. *macrophylla* or leaf tissue from *N*. *benthamiana* plants was ground using liquid nitrogen, and total RNA was extracted with the Trizol Reagent (Invitrogen) according to the procedure of the manufacturer. RNA extracts were resuspended in 100 µl of RNase-free water. RNA concentration was measured using a NanoDrop UV-Vis Spectrophotometer (Thermo Fisher Scientific). The RNA extracts were diluted with RNase-free water to adjust concentration to 1 ng/µl. Aliquots were stored at -80°C until use.

RT-PCR. Total RNA extracts were subjected to RT-PCR procedure using Phusion Flash High-Fidelity PCR Master Mix (Thermo Fisher Scientific) with a set of a forward primer (5'-CAGTGAATTCATGGACGACGAAACAAAGAAATTG-3') and a reverse primer (5'-CGATGGATCCTCAACGTGTGTTGAATTTCCCAAGC-3') annealing to the CP ORF of the CTV genome. The reactions were carried out according to the manufacturer. Reaction products were analyzed by electrophoresis in 1% agarose gels containing ethidium bromide at 200 ng/ml.

RT-qPCR. To determine the viral load in the inoculated plants, RT-qPCR was performed in a CFX Connect™ Real-time PCR detection system (BIO-RAD) using the iTaq™ Universal Probes One-Step kit (BIO-RAD) with 5 µl of iTaq universal probe reaction mix 2x (BIO-RAD), 2 µl RNAse-free water, 0.1 µl of iScript advanced reverse transcriptase (BIO-RAD), 0.4 µM of the forward (T36-RT-F) and reverse primers (T36-RT-R) covering the intergenic region between the RNA-dependent RNA polymerase and p33 ORFs of CTV (5-ACTTCGGACAAGCGGGTGAATT-3' and 5'-GCAAACATCTCGACTCAACTACC-3', respectively), 0.1 µM of 6-FAM/BHQ-1 labeled Taqman probe (5'-

AGCAACCGGCTGATCGATTGATT-3'), and 2 µl of total RNA extract (1ng/µl) in a total reaction volume of 10µl. Cycling conditions included reverse transcription at 50 °C for 5 min, incubation at 94 °C for 2 min, and 40 cycles of 94 °C for 10 s and 60°C for 40 s. Control samples in each run included total RNA from healthy citrus, water control, and at least two RNA transcript dilutions generated for the standard curve. Standard curve for the quantification of viral load in samples was created using *in-vitro* RNA transcripts, and serial dilutions were used in RT-qPCR assays to generate external standard curves as previously described (Bergua et al., 2016). Briefly, the cDNA molecules used as templates for the in vitro transcription were obtained by conventional RT-PCR using RNA extracts from the infected citrus bark with a set of a forward primer (T36-RT-F(T7)) and a reverse primer (T36-RT-R). The primer T36-RT-F(T7) is a modified version of T36-RF-F with the T7-promoter sequence at its 5´ end. PCR amplification product was in vitro transcribed using the MEGAscript® T7 Transcription Kit (Ambion, Inc.) following the manufacturer's instruction. Transcripts were treated with RNase-free DNase (Turbo DNase, Ambion, Inc.) and precipitated with 1.5X volume of DEPC-treated distilled water and 1.5X volume of 7.5M LiCl for 3 hours at -20 °C, collected by centrifugation at 14.000 rpm for 15 min at 4°C, washed with 70% ethanol and finally resuspended in DEPC-treated distilled water (with twice the volume of the transcription reaction). The concentration of the RNA transcripts was determined in duplicates using the Quant-iT^{TM} Ribogreen RNA Assay Kit (Invitrogen), and 10-fold serial dilutions containing 1010 to 101 RNA copies were prepared using total RNA extracts (10 ng/µl) from healthy *C. macrophylla* trees. The transcript RNA concentration (pmol) in each dilution was calculated using the following formula: µg of transcript RNA x (106 pg/1 µg) x (1 pmol/340 pg) x (1/number of bases of the transcript), and the number of RNA copies using this concentration value and Avogadro´s constant. Standard

curve was constructed plotting the threshold cycle (Ct) values from three independent assays with two replicates per standard dilution versus the logarithm of the RNA copy number.

Serological assays. Double-antibody sandwich ELISA of tissue extracts was performed as described previously[57] to confirm the presence of viruses. Briefly, 0.25 g of bark tissue from young flushes of *C*. *macrophylla* or leaf tissue from *N*. *benthamiana* plants was ground in 5 ml of the extraction buffer (3.2 mM Na_2HPO_4, 0.5 mM KH_2PO_4, 1.3 m KCL, 135 mM NaCl, 0.05% Tween® 20, pH 7.4) per sample. Purified IgG from rabbit polyclonal antiserum CTV-908 (1 ug/ml) was used as a coating antibody and a broadly reactive CTV monoclonal antibody ECTV172 was used for detection.

Virion analysis. 10 µl-drops of virion extracts were placed on the 400-mesh copper grids (Formvar/Carbon Square Mesh, UB, Electron Microscopy Sciences, USA) for negative contrast electron microscopy. After 5 min incubation, virion solution on each grid was removed using filter paper. The grids were stained using 1 % uranyl acetate dissolved in distilled water for 30 s followed by drying in the air. Images were taken using Hitachi H-7000 transmission electron microscope (Hitachi High-Technologies, Japan) equipped with a Veleta (2k X 2k) CCD side mount camera (EMSIS, Germany) at 100 kV. The length of virions was measured using Image J software (http://rsbweb.nih.gov/ij) directly from the images taken. For each mutant virus, over 100 flexuous virions were measured and up to 10 longest virions were selected to calculate the mean value.

Examination of fluorescence in plants. GFP fluorescence in the infiltrated leaves of *N*. *benthamiana* plants was observed using a hand-held UV lamp (365 nm, UVP, USA) set 6 inches

above the plants in a dark room. Bark tissue from citrus trees was examined for GFP fluorescence at different times beginning at six weeks after inoculation using Leica MZ10F fluorescence dissecting stereomicroscope with an attached Leica DFC310FX camera (Leica Microsystems) and processed using Leica LAS-X (Leica Microsystems) and Fiji software.

Confocal laser scanning microscopy. *N*. *benthamiana* leaves were cut into small squares and placed on glass slides with a drop of perfluorodecalin (Sigma-Aldrich) and observed using Leica TCS SP5 confocal laser scanning microscope system (Leica Micro- systems, Bannockburn, IL, USA). RFP was excited at 543 nm, and the emission was collected at 555 - 650 nm. CFP was excited at 405 nm, and the emission was collected at 430 - 480 nm. Images were captured using a 20X objective and processed using Leica LAS-X and Fiji software.

Statistical analysis. One way ANOVA followed by pairwise *post-hoc* Tukey's honest significant difference (HSD) test was conducted to determine the precise factor groups that differed. P values < 0.01 were regarded as statistically significant. Viral loads, the absolute number of gRNA copies per ng of total RNA extract from infiltrated *N*. *benthamiana*, were calculated as described earlier (Bergua et al., 2016). Exact n values for specific experiments were disclosed in each Figure.

References

Agranovsky, A.A., Lesemann, D.E., Maiss, E., Hull, R., Atabekov, J.G., 1995. 'Rattlesnake' structure of a filamentous plant RNA virus built of two capsid proteins. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 2470–2473.

Agranovsky, A.A., 1996. Principles of molecular organization, expression, and evolution of closteroviruses: Over the barriers. *Adv. Virus Res.* **47,** 119–158.

Agranovsky, A.A., 2016. Closteroviruses: Molecular biology, evolution and interactions with cells. In: *Plant Viruses: Evolution and Management*. 231-252 (Springer Singapore).

Alzhanova, D.V., Napuli, A.J., Creamer, R., Dolja, V.V., 2001. Cell-to-cell movement and assembly of a plant closterovirus: Roles for the capsid proteins and Hsp70 homolog. *EMBO J.* **20,** 6997–7007.

Alzhanova, D.V., Prokhnevsky, A.I., Peremyslov, V.V., Dolja, V.V., 2007. Virion tails of beet yellows virus: Coordinated assembly by three structural proteins. *Virology* **359,** 220–226.

Ambrós, S. *et al*., 2011. Agroinoculation of citrus tristeza virus causes systemic infection and symptoms in the presumed nonhost *Nicotiana benthamiana*. *Mol. Plant. Microbe. Interact.* **24,** 1119–1131.

Atallah, O.O. *et al*., 2016. A 5'-proximal region of the citrus tristeza virus genome encoding two leader proteases is involved in virus superinfection exclusion. *Virology* **489,** 108–115.

Bar-Joseph, M., Garnsey, S. M., Gonsalves, D., 1979. The closteroviruses: A distinct group of elongated plant viruses. *Adv. Virus Res.* **25,** 93–168.

Bergua, M., Kang, S.-H., Folimonova, S.Y., 2016. Understing superinfection exclusion by complex populations of citrus tristeza virus. *Virology* **499,** 331–339.

Boyko, V.P., Karasev, A.V., Agranovsky, A.A., Koonin, E.V. , Dolja, V.V., 1992. Coat protein gene duplication in a filamentous RNA virus of plants. *Proc. Natl. Acad. Sci. U. S. A.* **89,**

9156–9160.

Carrasco, P., Daròs, J.A., Agudelo-Romero, P., Elena, S.F., 2007. A real-time RT-PCR assay for quantifying the fitness of tobacco etch virus in competition experiments. *J. Virol. Methods* **139,** 181–188.

Cronshaw, J., Hoefert, L., Esau, K., 1966. Ultrastructural features of *beta* leaves infected with beet yellow virus. *J. Cell Biol.* **31**, 429–443.

Dolja, V.V, Karasev, A.V., Koonin, E.V., 1994. Molecular biology and evolution of closteroviruses: Sophisticated build-up of large RNA genomes. *Annu. Rev. Phytopathol.* **32,** 261–285.

Dolja, V.V., 2003. Beet yellows virus: The importance of being different. *Mol. Plant Pathol.* **4,** 91–98.

Dolja, V.V, Kreuze, J.F. , Valkonen, J.P.T., 2006. Comparative and functional genomics of closteroviruses. *Virus Res.* **117,** 38–51.

El-Mohtar, C., Dawson, W.O., 2014. Exploring the limits of vector construction based on citrus tristeza virus. *Virology* **448,** 274–283.

Febres, V.J. *et al*., 1996. The p27 protein is present at one end of citrus tristeza virus particles. *Phytopathology* **86,** 1331–1335.

Fernández-Calvino, L. *et al*., 2010. The helper-component protease transmission factor of tobacco etch potyvirus binds specifically to an aphid ribosomal protein homologous to the laminin receptor precursor. *J. Gen. Virol.* **91,** 2862–2873.

Forss, S., Schaller, H., 1982. A tandem repeat gene in a picornavirus. *Nucleic Acids Res.* **10,** 6441–6450.

Gorbalenya, A.E., Koonin, E.V., Lai, M.M., 1991. Putative papain-related thiol proteases of

positive-strand RNA viruses. Identification of rubi- and aphthovirus proteases and delineation of a novel conserved domain associated with proteases of rubi-, alpha- and coronaviruses. *FEBS Lett*. **288,** 201–205.

Gorbalenya, A.E., Enjuanes, L., Ziebuhr, J., Snijder, E.J., 2006. Nidovirales: Evolving the largest RNA virus genome. *Virus Res*. **117,** 17–37.

Harper, S.J., Cowell, S.J., Dawson, W.O., 2015. With a little help from my friends: Complementation as a survival strategy for viruses in a long-lived host system. *Virology* **478,** 123–128.

He, Y. *et al*., 2015. Deep sequencing reveals a novel closterovirus associated with wild rose leaf rosette disease. *Mol*. *Plant Pathol*. **16**, 449-458.

Hilf, M.E. *et al*., 1995. Characterization of citrus tristeza virus subgenomic RNAs in infected tissue. *Virology* **208,** 576–582.

Kambol, R., Kabat, P., Tristem, M., 2003. Complete nucleotide sequence of an endogenous retrovirus from the amphibian, Xenopus laevis. *Virology* **311,** 1–6.

Kang, S.-H., Bak, A., Kim, O.-K., Folimonova, S.Y., 2015. Membrane association of a nonconserved viral protein confers virus ability to extend its host range. *Virology* **482,** 208–217.

Karasev, A.V. *et al*., 1995. Complete sequence of the citrus tristeza virus RNA genome. *Virology* **208,** 511–520.

Karasev, A.V., 2000. Genetic diversity and evolution of closteroviruses. *Annu*. *Rev*. *Phytopathol* **38,** 293–324.

Kasschau, K.D., Cronin, S., Carrington, J.C., 1997. Genome amplification and long-distance movement functions associated with the central domain of tobacco etch potyvirus helper

component-proteinase. *Virology* **228,** 251–262.

Kasschau, K.D., Carrington, J.C., 1998. A counterdefensive strategy of plant viruses. *Cell* **95,** 461–470.

Liu, Y., Peremyslov, V.V, Medina, V., Dolja, V.V., 2009. Tandem leader proteases of Grapevine leafroll-associated virus-2: Host-specific functions in the infection cycle. *Virology* **383,** 291–299.

Lu, R. *et al*., 2004. Three distinct suppressors of RNA silencing encoded by a 20-kb viral RNA genome. *Proc*. *Natl*. *Acad*. *Sci*. *U*. *S*. *A*. **101,** 15742–15747.

Moreno, P., Ambrós, S., Albiach-Martí, M.R., Guerri, J., Peña, L., 2008. Citrus tristeza virus: A pathogen that changed the course of the citrus industry. *Molecular Plant Pathology* **9,** 251–268.

Napuli, A.J., Falk, B.W., Dolja, V.V., 2000. Interaction between HSP70 homolog and filamentous virions of the beet yellows virus. *Virology* **274,** 232–239.

Ng, J.C.K., Falk, B.W., 2006. Virus-vector interactions mediating nonpersistent and semipersistent transmission of plant viruses. *Annu*. *Rev*. *Phytopathol*. **44,** 183–212.

Pappu, H.R. *et al*., 1994. Nucleotide sequence and organization of eight 3' open reading frames of the citrus tristeza closterovirus genome. *Virology* **199,** 35–46.

Peng, C.-W., Dolja, V.V., 2000. Leader proteinase of the beet yellows closterovirus: Mutation analysis of the function in genome amplification. *J*. *Virol*. **74,** 9766–9770.

Peng, C.-W., Peremyslov, V.V, Mushegian, A.R., Dawson, W.O. , Dolja, V.V., 2001. Functional specialization and evolution of leader proteinases in the family closteroviridae. *J*. *Virol*. **75,** 12153–12160.

Peng, C.-W., Napuli, A. J., Dolja, V.V., 2003. Leader proteinase of beet yellows virus functions

in long-distance transport. *J*. *Virol*. **77,** 2843–2849.

Peremyslov, V.V, Hagiwara, Y., Dolja, V.V., 1998. Genes required for replication of the 15.5-kilobase RNA genome of a plant closterovirus. *J*. *Virol*. **72,** 5870–5876.

Robertson, C.J., Garnsey, S.M., Satyanarayana, T., Folimonova, S.Y., Dawson, W.O., 2005. Efficient infection of citrus plants with different cloned constructs of citrus tristeza virus amplified in *Nicotiana benthamiana* protoplasts, In: *16th IOCV Conference*. 187–195.

Santos, T.D.L., Botton, S.D.A., Weiblen, R., Grubman, M.J., 2006. The leader proteinase of Foot-and-Mouth disease virus inhibits the induction of *beta* interferon mRNA and blocks the host innate immune response. *J*. *Virol*. **80,** 1906–1914.

Satyanarayana, T. *et al*., 2000. Closterovirus encoded HSP70 homolog and p61 in addition to both coat proteins function in efficient virion assembly. *Virology* **278,** 253–265.

Satyanarayana, T., Gowda, S., Ayllón, M.A., Dawson, W.O., 2004. Closterovirus bipolar virion: Evidence for initiation of assembly by minor coat protein and its restriction to the genomic RNA 5' region. *Proc*. *Natl*. *Acad*. *Sci*. *U*. *S*. *A*. **101,** 799–804.

Simon-Loriere, E., Holmes, E.C., 2013. Gene duplication is infrequent in the recent evolutionary history of RNA viruses. *Mol*. *Biol*. *Evol*. **30,** 1263–1269.

Tatineni, S. *et al*., 2008. Three genes of citrus tristeza virus are dispensable for infection and movement throughout some varieties of citrus trees. *Virology* **376,** 297–307.

Tatineni, S., Gowda, S., Dawson, W.O., 2010. Heterologous minor coat proteins of citrus tristeza virus strains affect encapsidation, but the coexpression of HSP70h and p61 restores encapsidation to wild-type levels. *Virology* **402,** 262–270.

Tatineni, S., Robertson, C.J., Garnsey, S.M., Dawson, W.O., 2011. A plant virus evolved by acquiring multiple nonconserved genes to extend its host range. *Proc*. *Natl*. *Acad*. *Sci*. *U*.

S. A. **108,** 17366–17371.

Tian, T., Rubio, L., Yeh, H.H., Crawford, B., Falk, B.W., 1999. Lettuce infectious yellows virus: In vitro acquisition analysis using partially purified virions and the whitefly *Bemisia tabaci*. *J. Gen. Virol.* **80,** 1111–1117.

Tijms, M.A., Nedialkova, D.D., Zevenhoven-Dobbe, J.C., Gorbalenya, A.E., Snijder, E.J., 2007. Arterivirus subgenomic mRNA synthesis and virion biogenesis depend on the multifunctional nsp1 autoprotease. *J. Virol.* **81,** 10496–10505.

Tristem, M., Marshall, C., Karpas, A., Petrik, J., Hill, F., 1990. Origin of *vpx* in lentiviruses *Nature* **347,** 341-342.

Tzanetakis, I.E., Postman, J.D., Martin, R.R., 2005. Characterization of a novel member of the family closteroviridae from Mentha spp. *Phytopathology* **95,** 1043–1048.

Tzanetakis, I.E., Halgren, A., Mosier, N., Martin, R.R., 2007. Identification and characterization of raspberry mottle virus, a novel member of the Closteroviridae. *Virus Res.* **127,** 26–33.

Tzanetakis, I.E., Martin, R.R., 2007. Strawberry chlorotic fleck: Identification and characterization of a novel closterovirus associated with the disease. *Virus Res.* **124,** 88–94.

Valli, A., Martin-Hernández, A.M., López-Moya, J.J., García, J.A., 2006. RNA silencing suppression by a second copy of the P1 serine protease of Cucumber vein yellowingipomovirus,amemberofthefamilyPotyviridaethatlacksthecysteine protease HCPro. *J. Virol.* **80**, 10055–10063.

Valli, A., López-Moya, J.J., García, J.A., 2007. Recombination and gene duplication in the evolutionary diversification of P1 proteins in the family Potyviridae. *J. Gen. Virol.* **88,** 1016–1028.

Valli, A.A., Gallo, A., Rodamilans, B., López-Moya, J.J. , García, J.A., 2017. The HCPro from the Potyviridae family: an enviable multitasking Helper Component that every virus would like to have. *Molecular Plant Pathology* [Epub ahead of print] 10.1111/mpp.12553; 1-20.

Walker, P.J. *et al*., 1992. The genome of bovine ephemeral fever rhabdovirus contains two related glycoprotein genes. *Virology* **191,** 49–61.

Willemsen, A., Zwart, M.P., Higueras, P., Sardanyés, J., Elena, S.F., 2016. Predicting the stability of homologous gene duplications in a plant RNA virus. *Genome Biol*. *Evol*. **8,** 3065–3082.

Zhu, H. Y., Ling, K.S., Goszczynski, D.E., McFerson, J.R., Gonsalves, D., 1998. Nucleotide sequence and genome organization of grapevine leafroll-associated virus-2 are similar to beet yellows virus, the closterovirus type member. *J*. *Gen*. *Virol*. **79,** 1289–1298.

Ziebuhr, J., Snijder, E.J., Gorbalenya, A.E., 2000. Virus-encoded proteinases and proteolytic processing in the Nidovirales . *J*. *Gen*. *Virol*. **81,** 853–879.

Acknowledgements

We thank Yodiany X. Arroyo for the assistance with plant maintenance. This research was supported by the National Science Foundation under Grant Number 1615723 (to S. Y. Folimonova).

Author contributions

S.Y.F. conceived the study. S.-H.K., O.O.A., and S.Y.F. designed the research. S.-H.K., O.O.A., Y.-D.S., and S.Y.F. performed the research. S.-H.K., O.O.A., Y.-D.S., and S.Y.F. analyzed the data. S.-H.K. and S.Y.F. wrote the manuscript with contributions from all the authors. S.-H.K. and O.O.A. contributed equally to this work.

Figure Legends

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.

Fig. 2. Virus accumulation in inoculated leaves of *N. benthamiana*. Images show GFP fluorescence observed from *N. benthamiana* plants in which CTV variants were introduced using *Agrobacterium tumefaciens*-mediated infiltration at six days post infiltration (dpi) under a hand-held UV light **(A)** or fluorescence microscope **(B)**. No detectable GFP fluorescence was observed from *N. benthamiana* infiltrated with the ΔL1 virus with a hand-held UV light. Bars = 100 μm. **(C)** Accumulation of viruses in *N. benthamiana* plants [viral load = number of genomic RNA copies per ng of total RNA]. Each bar represents a mean of six independent biological replicates ($N = 6$; two sets of samples consisting of one leaf per plant from two plants infiltrated with each virus variant were collected at three independent agro-inoculations and total RNA extracted from six sets of samples per each virus variant). The absolute number of CTV RNA

copies was determined by RT-qPCR based on a standard curve generated by amplification of known amounts of the target under conditions identical to those of the experimental sample as described in Methods. For each set of leaves, two RT-qPCR determinations were made. Bars represent mean values ± standard error of the mean.

Fig. 3. Accumulation of mutant viruses with alterations within the L1 coding region in systemic leaves of *N. benthamiana* and citrus. (A) Images show GFP fluorescence observed in systemic leaves of *N*. *benthamiana* plants in which CTV variants were introduced using *A*. *tumefaciens*-mediated infiltration. Images were taken at six weeks post infiltration (wpi) under a hand-held UV light. No detectable GFP fluorescence was observed from *N*. *benthamiana* infiltrated with ΔL1 and BT-L1NTD. **(B)** RT-PCR analysis of virus accumulation. Total RNA obtained from systemic leaves of all biological replicates (described in Fig. 2C) from the same mutant virus was combined for RT-PCR. The size of the expected amplification product with a set of primers annealing to the CP coding region was 800 bp. **(C)** Analysis of virus titer in plants shown in (A) using ELISA with the CTV-specific antibody. Each bar represents a mean of the virus titer (A_{405} values of ELISA) ± standard error of the mean from six independent biological replicates tested in two separate assays (N = 12). **(D)** Images show GFP fluorescence observed in phloem-associated cells of *C*. *macrophylla* upon virus inoculation at three months after inoculation. Observations were made on the internal surface of bark using a dissecting fluorescence microscope. Bars = 0.3 mm. **(E)** RT-PCR analysis of virus accumulation. Total RNA obtained from newly developed flushes of all biological replicates (independently inoculated citrus trees, N = 10-12) from the same mutant virus was combined for RT-PCR. The expected amplification product with a set of primers annealing to the CP coding region was 800

bp. Note that none was loaded in the lane labeled as ΔL1 (*). **(F)** Analysis of virus titer in plants shown in (D) using ELISA with the CTV-specific antibody. Each bar represents a mean of the virus titer (A_{405} values of ELISA) ± standard error of the mean of samples described in (E) tested in two separate assays.

Fig. 4. Accumulation of mutant viruses with alterations within the L2 coding region in systemic leaves of *N. benthamiana* and citrus. (A) Images show GFP fluorescence observed in systemic leaves of *N*. *benthamiana* plants in which CTV variants were introduced via *A*. *tumefaciens*-mediated infiltration. Images were taken at six wpi under UV light. The image for WT is the same as that used in Fig. 3A. **(B)** Comparison of GFP fluorescence observed in the systemic leaves infected by WT or ΔL2. **(C)** RT-PCR analysis of virus accumulation. Total RNA obtained from all biological replicates (described in Fig. 2C) of the same mutant virus was combined for RT-PCR. The size of the expected amplification product with a set of primers annealing to the CP coding region was 800 bp. **(D)** Analysis of virus titer in plants shown in (A) using ELISA with the CTV-specific antibody. Each bar represents a mean of the virus titer (A_{405} values of ELISA) ± standard error of the mean from six independent biological replicates tested in two separate assays (N = 12). **(E)** Comparison of symptoms observed in the systemic leaves infected by WT or ΔL2. **(F)** Symptoms produced by plants infected by WT or ΔL2. **(G)** Images show GFP fluorescence observed in phloem-associated cells of *C*. *macrophylla* upon inoculation with each virus at 3 months after inoculation. Observations were made on the internal surface of bark using a dissecting fluorescence microscope. The image for WT is the same as that used in Fig. 3d. Bars = 0.3 mm. **(H)** RT-PCR analysis of virus accumulation. Total RNA obtained from all biological replicates (independently inoculated citrus trees, N = 10-12) of the same mutant

virus was combined for RT-PCR. The size of the expected amplification product with a set of primers annealing to the CP coding region was 800 bp. **(I)** Analysis of virus titer in plants shown in (G) using ELISA with the CTV-specific antibody. Each bar represents a mean of the virus titer (A_{405} values of ELISA) ± standard error of the mean of samples described in (H) tested in two separate assays.

Fig. 5. *In vivo* characterization of the virions produced by mutant viruses carrying alterations in the L1 or L2 coding sequences. (A) Negative contrast electron microscopy of virions produced by the WT, *Δ*L2, BT-L1NTD, BT-L2NTD, BT-L1Pro, and BT-L2Pro viruses. Virions were purified from *N. benthamiana* leaves shown in Fig. 2A and examined using TEM. Representative images are displayed. Bars = 100 nm. **(B)** A mean length of virions calculated based on the TEM images of ten longest virions per treatment ± standard error of the mean (*N* = 10).

Fig. 6. Subcellular localization of CTV proteases. (A) Images show single sections of *N. benthamiana* leaves infiltrated with CFP:L1 or RFP:L2. Panels on the right are images merged with the bright field. White arrows indicate punctate spots along the cellular peripheral membrane. Bars = 20 μm. **(B)** Images show the motility of the fluorescing body formed by RFP:L2 inside the cell in representative time-lapse micrographs (14 seconds in 2 seconds interval) extracted from Supplementary video S3. White arrows follow the motile object. **(C)** Hydrophobicity plot (shown in blue) and transmembrane domain (shown in the grey box in the middle of L1) were analyzed using MPEx (http://blanco.bio.mol.uci.edu/mpex/) and TMpred (http://www.ch.embnet.org/software/TMPRED_form.html/).

Figure 1

Click here to download high resolution image

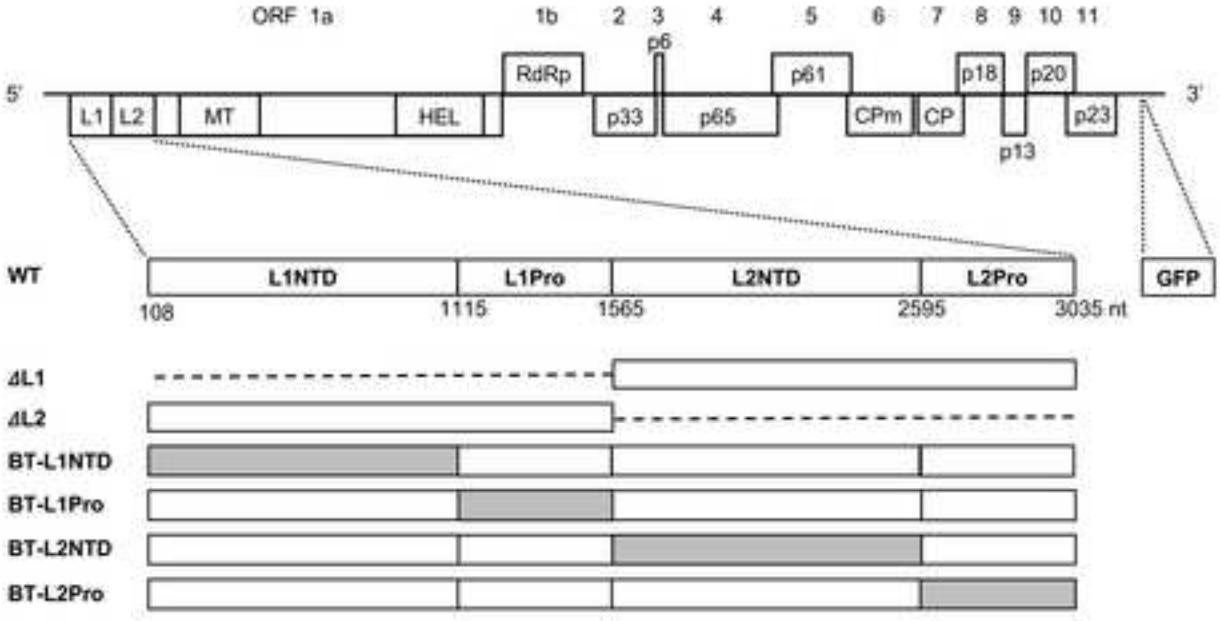

Figure 2

Click here to download high resolution image

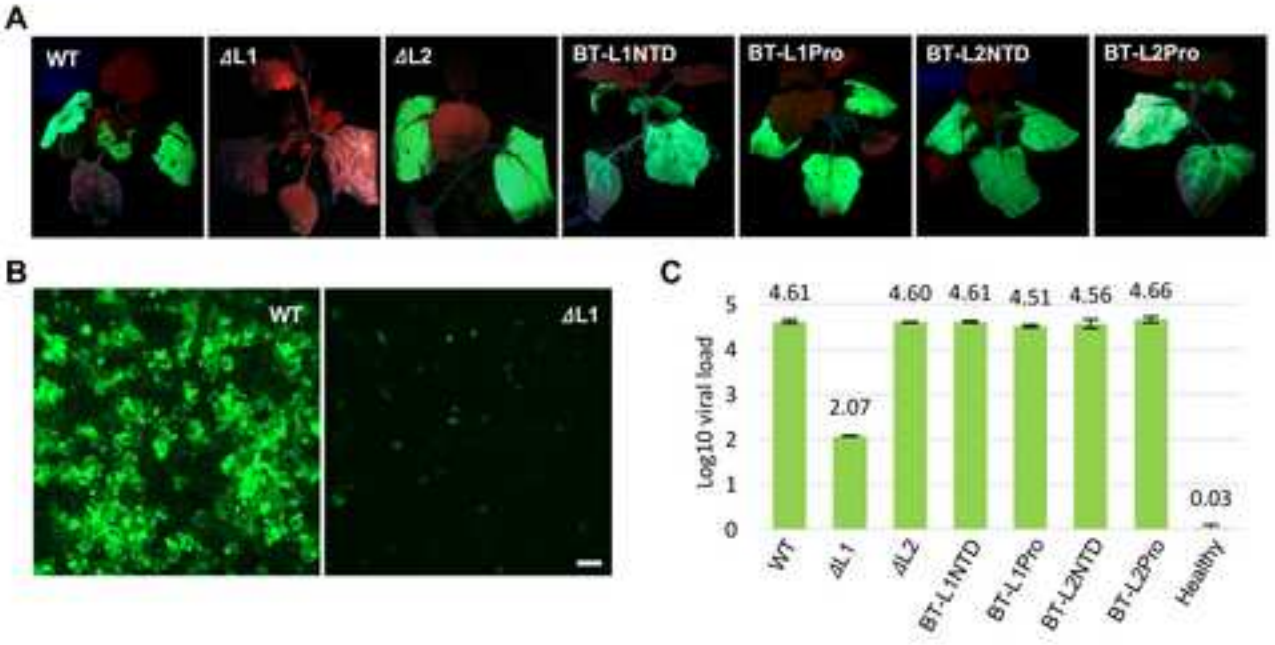

Figure 3

Click here to download high resolution image

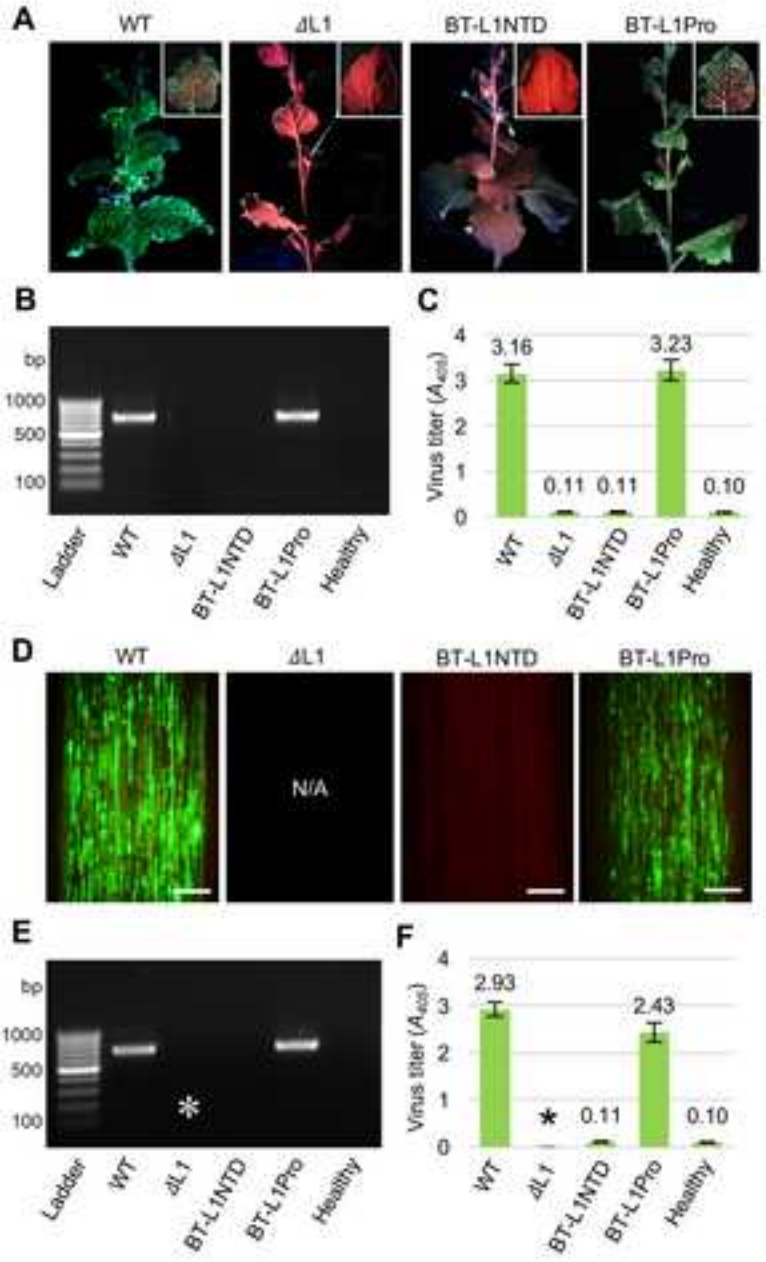

Figure 4

Click here to download high resolution image

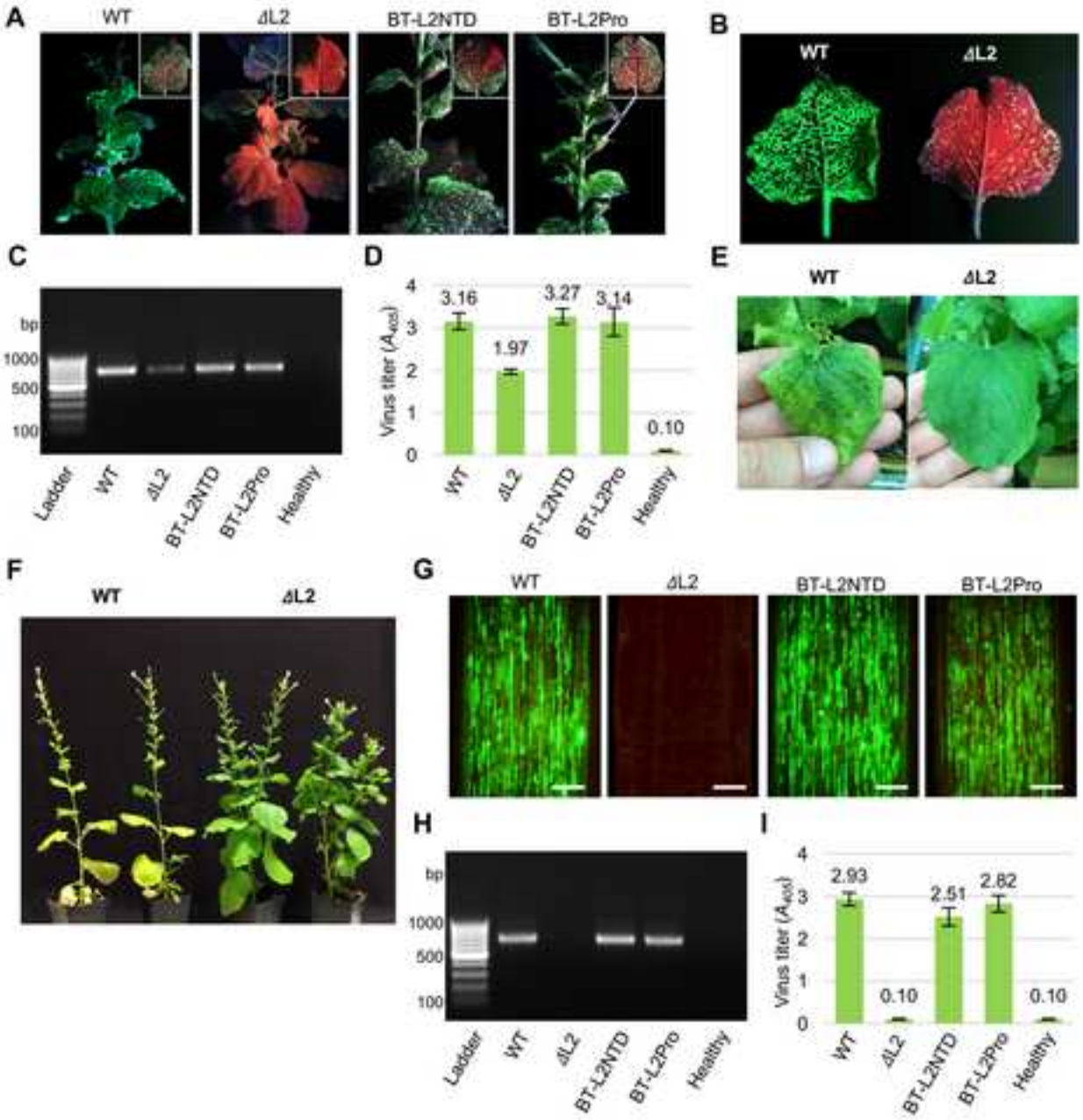

Figure 5

Click here to download high resolution image

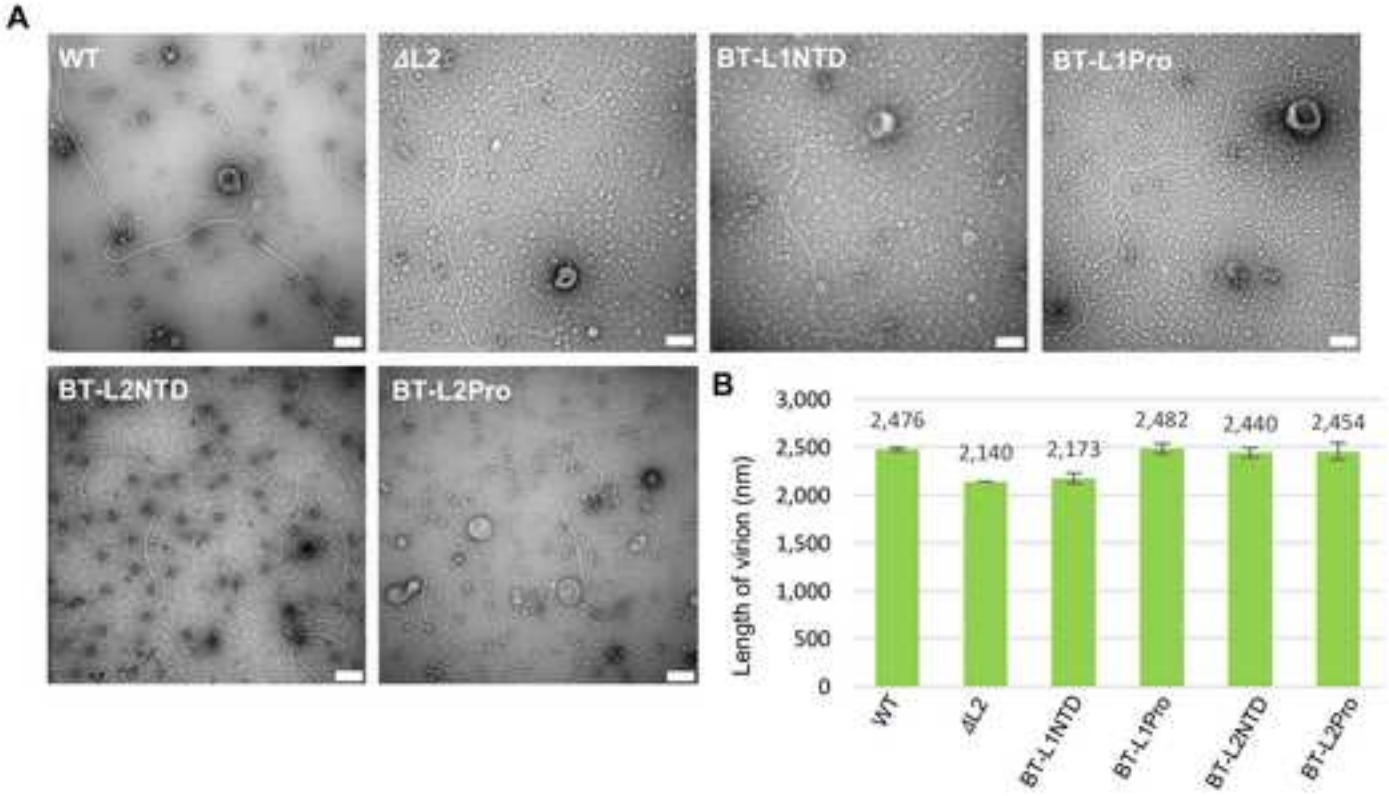

Figure 6

Click here to download high resolution image