Genome sequence of a divergent Colombian isolate of potato virus V (PVV) infecting Solanum phureja

P. GUTIÉRREZ¹, H. JARAMILLO MESA¹, M. MARÍN MONTOYA²

¹Laboratorio de Microbiología Industrial, Facultad de Ciencias, Universidad Nacional de Colombia Sede Medellín, Calle 59A No 63 - 20 Medellín, Colombia; ²Laboratorio de Biología Celular y Molecular, Facultad de Ciencias, Universidad Nacional de Colombia Sede Medellín, Calle 59A No 63 - 20 Medellín, Colombia

Received March 6, 2015; revised June 2015; accepted February 2, 2016

Summary. – Deep sequencing analysis of the transcriptome of a Solanum phureja cv. Criolla Colombia plant with symptoms typical of a virus disease revealed an infection with potato virus V (PVV). The PVV-phureja genome comprises 9904 nt, exhibits 83% nucleotide identity with currently fully sequenced PVV isolates and contains one large ORF that codes for a polyprotein of 3065 residues flanked by 5' and 3' UTR of 217 and 448 nt, respectively. Phylogenetic analysis of the PVV-phureja polyprotein indicates that it is divergent with respect to most PVV isolates. This is the first complete PVV genome of an isolate infecting a host different to S. tuberosum and, to this date, the only one from the South American Andes.

Keywords: potato virus V; potyvirus; Solanum phureja; next generation sequencing

Introduction

Potato virus V (PVV) is a member of the genus Potyvirus (the family Potyviridae) first characterized by Rozendaal et al. (1971) and initially considered to be a deviating strain of potato virus Y (PVY). Further serological and phylogenetic studies, however, confirmed it to be a distinct potyvirus species clearly distinguishable from PVY (Fribourg and Nakashima, 1984; Oruetxebarria et al., 2000; Inoue-Nagata et al., 2002; Spetz et al., 2003). Symptoms induced by PVV are mild and are characterized by leaf pallor, some decrease in leaflet size and slight leaflet distortion; in some cases, PVV can also induce the development of mosaics and necrotic spots in lower leaves (Calvert et al., 1980; Fribourg and Nakashima, 1984; Jones and Fuller, 1984; Jones, 1987). PVV can be transmitted by infected seed tubers and non-persistently by aphids such as Myzus persicae, Brachycaudus helichrysi, Macrosiphum euphorbiae and Rhopalosiphoninus latysiphon (Calvert et al., 1980; Bell, 1983; Fribourg and Nakashima, 1984). Potato virus V has been reported to infect potato cultivars in the Andean region of South America (Spetz et al., 2003), Northwestern Europe (Oruetxebarria et al., 2000) and, more recently, Iran (Shamsadden-Saeed et al., 2014). This limited distribution is apparently due to the presence of hypersensitive resistance genes and broad-spectrum resistance in potato cultivars (Barker 1997; Valkonen 2008).

The first evidence of PVV infecting potato cultivars in Colombia came from the analysis of a root RNA-virome of a Solanum phureja cv. Criolla Colombia plant. This work revealed the presence of a PVV variant with 90% amino acid identity in the coat protein region to currently known PVV sequences suggesting the existence of a novel isolate infecting a host different to S. tuberosum (Gutiérrez et al., 2014). To date, only two PVV genomes have been fully sequenced: DV42 from Scotland (Oruetxebarria et al., 2000) and KER. LAL.P from Iran (Shamsadden-Saeed et al., 2014) and, in spite of being considered the center of origin of PVV (Spetz
et al., 2003), there are no complete genomes of PVV isolates from the South American Andes. To better characterize the PVV variant infecting *S. phureja* in Colombia, we have sequenced the complete genome of an isolate derived from an infected plant showing mild mosaics and discuss its phylogenetic relation to other PVV isolates.

**Materials and Methods**

RNA extraction and next generation sequencing. This work was performed on a bulk leaflet sample obtained from an *S. phureja* var. Criolla Colombia plant collected in the municipality of la Unión (Antioquia, Colombia). Total RNA was extracted with the GeneJET Plant RNA Purification Mini Kit (Thermo Fisher Scientific) using fresh leaflet tissue ground with liquid nitrogen. The library was constructed with the TrueSeq RNA Sample Preparation Kit (Illumina); rRNA was depleted with The TruSeq Stranded Total RNA with Ribo-Zero Plant kit (Illumina). Sequencing was performed with the Illumina HiSeq 2000 system provided by Macrogen (South Korea), with an average coverage depth of 76,013X. ORFs codifying for viral proteins were assembled from a total of 7,539,884 paired-end reads with an average length of 4,797,325,400 bp. Adapter sequences and low quality bases were removed with SeqTK (https://github.com/lh3/seqtk).

Bioinformatic analysis. The PVV-phureja genome was assembled by de novo reconstruction of the *S. phureja* transcriptome with Trinity (Grabherr et al., 2011) and confirmed by mapping the reads with Bowtie2 (Langmead and Salzberg, 2012). The final contig was assembled from a total of 7,539,884 paired-end reads with an average coverage depth of 76,013X. ORFs codifying for viral proteins were identified with BLASTX (Gish and States, 1993). Sequence alignments were performed with MUSCLE (Edgar, 2004) and phylogenetic trees were obtained with the neighbor joining method (Tamura et al., 2011). The assembled sequence was deposited in GenBank under Acc. No. KP849483 with PVV-phureja as isolate name.

Table 1. Percent amino acid sequence identities of the individual proteins of PVV-phureja with those of its closest relatives within the genus *Potyvirus*

<table>
<thead>
<tr>
<th>Protein</th>
<th>P1</th>
<th>HC-Pro</th>
<th>P3</th>
<th>6K1</th>
<th>CI</th>
<th>6K2</th>
<th>VPg</th>
<th>Nla-Pro</th>
<th>Nlb</th>
<th>CP</th>
</tr>
</thead>
<tbody>
<tr>
<td>PVV KER</td>
<td>77.85</td>
<td>83.77</td>
<td>89.66</td>
<td>94.23</td>
<td>96.69</td>
<td>88.46</td>
<td>90.43</td>
<td>90.65</td>
<td>90.94</td>
<td>95.59</td>
</tr>
<tr>
<td>PVV DV42</td>
<td>70.59</td>
<td>84.43</td>
<td>87.15</td>
<td>94.12</td>
<td>96.53</td>
<td>90.38</td>
<td>94.68</td>
<td>90.65</td>
<td>94.22</td>
<td>93.38</td>
</tr>
<tr>
<td>WPMV</td>
<td>54.67</td>
<td>80.92</td>
<td>71.43</td>
<td>92.31</td>
<td>91.32</td>
<td>75.00</td>
<td>88.30</td>
<td>83.74</td>
<td>87.48</td>
<td>86.76</td>
</tr>
<tr>
<td>PTMV</td>
<td>51.56</td>
<td>80.92</td>
<td>60.50</td>
<td>90.38</td>
<td>91.17</td>
<td>72.55</td>
<td>76.06</td>
<td>83.74</td>
<td>86.71</td>
<td>90.44</td>
</tr>
<tr>
<td>PmoV</td>
<td>36.33</td>
<td>65.21</td>
<td>43.23</td>
<td>78.43</td>
<td>76.03</td>
<td>65.38</td>
<td>78.72</td>
<td>70.73</td>
<td>75.53</td>
<td>73.43</td>
</tr>
<tr>
<td>BSMV</td>
<td>33.67</td>
<td>64.99</td>
<td>45.51</td>
<td>70.59</td>
<td>73.82</td>
<td>55.77</td>
<td>69.68</td>
<td>67.89</td>
<td>75.14</td>
<td>76.21</td>
</tr>
<tr>
<td>VVY</td>
<td>36.73</td>
<td>62.58</td>
<td>33.72</td>
<td>77.08</td>
<td>69.72</td>
<td>50.00</td>
<td>78.19</td>
<td>70.73</td>
<td>75.53</td>
<td>73.43</td>
</tr>
<tr>
<td>BMV</td>
<td>39.72</td>
<td>62.96</td>
<td>37.46</td>
<td>58.82</td>
<td>70.50</td>
<td>55.77</td>
<td>71.81</td>
<td>65.85</td>
<td>71.24</td>
<td>73.61</td>
</tr>
<tr>
<td>PVY</td>
<td>35.29</td>
<td>62.50</td>
<td>37.68</td>
<td>70.78</td>
<td>70.03</td>
<td>46.70</td>
<td>71.81</td>
<td>61.38</td>
<td>72.06</td>
<td>72.96</td>
</tr>
</tbody>
</table>

Results and Discussion

**Genome sequence properties**

The assembled PVV-phureja genome has 9904 nt [excluding the 3' poly(A) tract] and is 83% identical at the whole genome level to isolates DV42 and KER.LAL.P, the only complete PVV genomes available at the moment. The PVV-phureja genome contains one large ORF that codes for a polyprotein of 3065 residues flanked by 5' and 3' UTR of 217 and 448 nt, respectively. The 5' UTR contains the sequences ACAAT ACATAAC (positions 24–35) and GAAAACGA (positions 46–52) that correspond to potyboxes A and B (Turpen, 1989). Putative polyprotein cleavage sites were identified at amino acid positions 289, 745, 1102, 1788, 1840, 2028, 2274 and 2793 that give rise to ten mature proteins: P1, HC-Pro, P3, 6K1, CI, 6K2, Nla-VPg, Nla-Pro, Nlb and CP. Table 1 presents the sequence identity for each mature protein with respect to selected members of the genus *Potyvirus*.

Proteinase P1 comprises 289 residues (32.47 kDa) and is cleaved at the C-terminal end at sequence MVQF/S, which is identical in PVV isolates DV42, and KER.LAL.P (Adams et al., 2005). P1 is the least conserved protein in PVV-phureja sharing 70–77% amino acid identity with homologs within the PVV group in agreement with previous observations, which showed P1 to be the protein most variable in size and sequence within the genus *Potyvirus* (Rohozková and Navrátil, 2011). A recent study in tobacco etch virus (TEV) has shown P1 to contain nucleolar localization and nuclear export signals that allow its traffic in and out of the nucleus and nucleolus during infection (Martínez and Daròs, 2014). Sequence alignment with PVV-phureja suggests that the nucleolar localization signal is placed at positions 73–95 (LRKASERKVRKSKEL IAARRREL), while the nuclear export signals maps at residues 235–244 (LRRGDSGIIIL). The first motif has a moderate

NCBI Acc. Nos: PVV KER (potato virus V KER.LAL.P: AGT56190); PVV DV42 (potato virus V DV42: NP_659008); WPMV (wild potato mosaic virus: NP_741959); PTMV (Peru tomato mosaic virus: NP_787937); PmoV (pepper mottle virus: NP_041276); BSMV (Brugmansia suaveolens mottle virus: YP_003900466); VVY (Verbena virus Y: YP_001931955); BMV (Bidens mosaic virus: YP_008877631); PVY (potato virus Y: AAW79560).
The helper component proteinase HC-Pro corresponds to a polypeptide of 456 aa (51.91 kDa), cleaved at sequence YRVG/G in agreement with the classical motif YXVG/G; as with P1 protease, cleavage sites are identical to those predicted in PVV isolates DV42 and KER.LAL.P. The conserved KITC motif in Domain I (N-terminus) has been found to be essential for aphid transmission (Peng et al., 1998). However in PVV-phureja, the Ile residue is replaced by Leu, this KLC (residues 49–52) variant is also observed in potyviruses closely related to PVV such as Peru tomato mosaic virus (PTV), wild potato mosaic virus (WPMV) and pepper mottle virus (PepMoV). The PKT motif (residues 308–310), located in the central part of the protein (Domain II), is also involved in aphid transmission and replication-maintenance (Peng et al., 1998). It contains FRNK (residues 179–182) and IGN (residues 248–250) motifs involved in RNA silencing suppressor (RSS) activity and the amplification process, respectively (Kasschau et al., 1997; Shiboleth et al., 2007). A CCC sequence motif (residues 290–292) necessary for the systemic spread of the virus within the host is also present (Kasschau et al., 1997). Finally, domain III has a cysteine-type proteinase motif characterized by a GYCY box (residues 340–343) that catalyzes the autoproteolytic cleavage from the polyprotein (Guo et al., 2011).

The P3 protein in PVV-phureja is composed of 357 residues (40.99 kDa); this protein is required for viral replication (Klein, et al., 1994) but its precise role remains obscure. Recent experiments showed that P3 is targeted to the membranes of the endoplasmic reticulum (ER) and forms inclusions associated with the Golgi apparatus that traffic along the actin filaments and colocalize with replication vesicles (Revers and Garcia, 2015). Two hydrophobic regions were identified in the P3 of several potyviruses, and the one located in the C-terminal end of the protein was shown to be responsible for the ER targeting of P3 (Cui et al., 2010). This transmembrane domain is predicted to be located between residues 261 and 281. Slippage product P3N-PIPO is predicted to be a basic polypeptide of 231 residues (26.0 kDa) that results from a frameshift at sequence GA,C (2,938–2,946) within the P3 ORF and ends with a UGA stop codon at position 3,145.

The 6K1 peptide of PVV-phureja comprises 52 residues (5863.8 kDa) and shares 94% amino acid identity with respect to available PVV sequences. Interestingly, all 6K1 proteins from PVV isolates sequenced to date (GenBank Acc. Nos: NP_734372, AGT56190, NP_659008) are 100% identical; in PVV-phureja the following substitutions were observed: S4A, I13V and S17T.

Cylindrical inclusion (CI) protein consists of 634 residues (79.94 kDa) and is the best conserved polypeptide in PVV-phureja with respect to other PVV isolates (96% identity). CI contains the DEXH (DECH, residues 174–177) and GSGKSX3P (GSGKSTGLP, residues 88–96) motifs typical of SF2 superfamily of helicases (Fernández et al., 1997). Helicase ATP-binding (residues 72–224) and Helicase C-terminal (residues 243–402) domains are also present in the protein.

6K2 is a 52-residue protein (5675.6 kDa). In turnip mosaic virus (TuMV), 6K2 induces the formation of vesicles involved in replication and intracellular movement of viral RNA (Grangeon et al., 2013). With respect to PVV isolates, PVV-phureja has the following unique mutations: A4T, G7T, I14L, A20V and T43V.

The first nuclear inclusion protein, NIa, possesses two domains: VPg, a viral genome-linked protein, and NIa-Pro, a protease responsible for cleavage of most proteins of the precursor polyprotein (Adams et al., 2005). The viral genome-linked protein VPg is a protein involved in...
multiple steps of the potyvirus infection cycle, including viral multiplication and movement (Rajamäki et al., 2014). The corresponding protein in PVV-phureja comprises 188 amino acids (21.57 kDa) and shares 90.43–94.68% aa identity with respect to isolates KER.LAL.P and DV42, respectively. Nla-Pro is a cysteine protease responsible for cleavage at the P3/6K1, 6K1/CI, CI/6K2, 6K2/VPg, VPg/Nla-Pro, Nla-Pro/Nib and Nib/CP junctions of the potyvirus polyprotein. Nla-Pro is a 246 residue protein (28.04 kDa) cleaving at consensus sequence V-X-(HE)-(QE) identical to cleavage sites in isolates DV42 and KER.LAL.P (Fig. 1).

The RNA-dependent RNA polymerase Nlb (59.74 kDa) contains the RdRp motif GDD and nuclear localization signals NSLI and NSLII. NSLI consists of residues 1–16 with sequence GRSTGWMLNALKENLQ, which is 56% similar to the NLSI motif in TEV (Li et al., 1997); NSLII, on the other hand, comprises residues 292–316 (TPISTPDGTILKKFRGNNSGQPSTV) with only three substitutions with respect to the TEV consensus (L295S, H305F and K306R).

Finally, the CP protein (30.45 kDa) in PVV-phureja consists of variable N- and C-terminal domains and a conserved core region involved in RNA binding (López-Moya et al., 1999). CP shares 100% aa identity to the CP protein of PVV-Colombia, an isolate previously identified in root transcriptome analysis of an infected S. phureja plant from the municipality of La Unión (Antioquia) (Gutiérrez et al., 2014); with respect to available CP sequences, PVV-phureja shares 93–96 percent identity at the amino acid level with PVV isolates infecting S. tuberosum. The highly conserved DAG motif involved in aphid transmission (López-Moya et al., 1999) is located at positions 9–11. Eight amino acids substitutions unique to PVV-phureja or rare in other PVV isolates were observed: A5T, S7T, A24A, A26S, L34V, A41T, I49V, V64A, 1154V, N182S.

![Neighbor joining tree showing the phylogenetic relationship of PVV-phureja to viruses of the genus Potyvirus using the complete polyprotein (a) and coat protein region (b) sequences](image-url)

Trees are drawn to scale, with branch lengths measured in the number of substitutions per site. All positions containing gaps and missing data were eliminated. Bootstrap values > 50% are indicated at each node. GenBank (gb) Acc. Nos. are shown in parentheses.
Phylogenetic analysis

The complete PVV-phureja polyprotein clusters within a group that includes PVV DV42 (Scotland), PVV KER, LAL.P (Iran), and a sister group including WPMV and PTV. From this phylogenetic tree it can be concluded that PVV-phureja has a closer affinity to WPMV and PTV than its Eurasian counterparts (Fig. 2a). A similar analysis using only the coat protein clearly shows that PVV-phureja is a member of the PVV group. This analysis separates existing PVV sequences into three main groups according to the host they infect: S. tuberosum, S. phureja and S. tuberosum cv. Papa Amarillo. The S. tuberosum group includes all known PVV isolates from Europe, Iran and Peru. The clade infecting Solanum tuberosum cv. Papa Amarillo includes isolate PA-4, which is a serologically distinct strain of PVV that cannot be detected by commonly used antisera (Shiel et al., 2004). As expected, the coat protein of PVV-phureja clusters with the PVV-Colombia isolate (mentioned above). PVV isolate PA-4 branches at the most basal node separating it from the S. tuberosum and S. phureja groups. This could suggest that the ancestral potyvirus of the PVV group first diverged into other species of the genus Solanum, before giving rise to the best-studied species infecting S. tuberosum.

S. phureja cv. Criolla is a diploid potato species that produces tubers with yellow skin and flesh widely grown in the Andes from western Venezuela to the center of Bolivia (Ghislain et al., 2006). This crop has received increasing attention in Colombia as a potential exotic export product due to its excellent organoleptic properties; high levels of vitamins A, B, C; short production cycle and reduced tuber dormancy (Rodríguez et al., 2009). S. phureja represents about 10% of the total 160,000 ha dedicated to potato in Colombia and one of the most limiting factors for its production is the high susceptibility to viral infections (MADR 2005). In spite of barely producing symptoms, PVV could be an important pathogen for S. phureja in Colombia as evidenced by the high levels of viral RNA found in the infected plant (31.43% of reads). This work suggests that control measures must be implemented to restrict the spread of PVV-phureja into other regions and probably should be included in current tuber seed-certification programs. Further studies are in progress to investigate its host range, transmission mode, synergistic interactions with other viruses and overall agronomical effects. To our knowledge, this is the first complete genome of an Andean PVV strain and also the first from a host different to S. tuberosum.

Acknowledgements. This work was funded by Universidad Nacional de Colombia (Grant VRI. 19438) and International Foundation for Science (Sweden, Grant: C/4634-2).

References


