High-throughput sequencing of *Potato virus M* from tomato in Slovakia reveals a divergent variant of the virus

MIROSLAV GLASA¹*, KATARÍNA ŠOLTYS²,³, LUKÁŠ PREDAJŇA¹, NINA SIHELSKÁ¹, JAROSLAV BUDIŠ⁴, MICHAELA MRKVOVÁ⁵, JÁN KRAIC⁵,⁶, DANIEL MIHÁLIK⁵,⁶,⁷, ANA BELÉN RUÍZ-GARCÍA⁸

¹Institute of Virology, Biomedical Research Centre, Slovak Academy of Sciences, Bratislava, Slovak Republic; ²Comenius University Science Park and ³Faculty of Natural Sciences, Comenius University in Bratislava, Bratislava, Slovak Republic; ⁴Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava, Mlynská dolina, Slovak Republic; ⁵Department of Biotechnologies, Faculty of Natural Sciences, University of SS. Cyril and Methodius, Trnava, Slovak Republic; ⁶National Agriculture and Food Centre, Research Institute of Plant Production, Piešťany, Slovak Republic; ⁷Institute of High Mountain Biology, University of Žilina, Žilina, Slovak Republic; ⁸Instituto Valenciano de Investigaciones Agrarias, Center of Plant Protection and Biotechnology, Moncada, Spain

*Corresponding author: miroslav.glasa@savba.sk


Abstract: High-throughput sequencing (HTS) analysis of tomato (*Solanum lycopersicum*) samples revealed the presence of *Potato virus M* (*PVM*) in this crop in Slovakia. Full-length genomes of three PVM isolates were obtained using both HTS and Sanger sequencing validation. While two isolates (T40 and T50) were shown to belong to major Group I, a divergent T20 isolate was phylogenetically unrelated to any known PVM variant, potentially representing a new phylogenetic group. Despite a relatively high intraspecies diversity (17.3 ± 0.3%), no evidence of recombination was detected in the dataset of available complete PVM sequences. Conventional screening of tomato plants in Slovakia using ELISA and RT-PCR further confirmed a frequent occurrence of PVM in this host. Developed RT-PCR showed its polyvalence to detect the PVM Group I isolates, however, in silico analysis of primer binding sites indicated its compromised use for Group II isolates. Our results further pinpoint the significance of HTS for unbiased unveiling of virus diversity and a need for continual optimisation of molecular detection tools.

Keywords: PVM; *Solanum lycopersicum* L.; Carlavirus; full-length genome; RT-PCR; phylogenetical diversity

Tomato (*Solanum lycopersicum* L.) represents an economically attractive agricultural crop in Slovakia due to its nutritional properties and usefulness for industrial processing. However, viral pathogens remain an important factor potentially limiting tomato yield and quality (Hanssen et al. 2010; Jones et al. 2014).

Knowledge of viral genetic diversity and the understanding of its complexity are crucial for efficient management of viral disease, i.e. through development...
MATERIAL AND METHODS

Determination of complete genome sequences of PVM isolates. Total RNAs from tomato leaves were extracted using the NucleoSpin RNA Plant kit (Macherey-Nagel, Duren, Germany) and ribosomal RNA was removed using the Ribo-Zero rRNA Removal Kit (Illumina, San Diego, USA). The rRNA-depleted total RNAs sample was used for double-stranded cDNA synthesis using the SuperScript II kit (Thermo Fisher Scientific, Waltham, USA). The column-purified cDNA (DNA Clean & Concentrator™-5 – DNA kit; Zymo Research, Irvine, USA) was then processed with the transposon-based chemistry library preparation kit (Nextera XT; Illumina, USA). Low-cycle PCR and mutual indexing of the fragments were carried out. Fragments were purified with 1.8x AMPure XP beads (BeckmanCoulter, Indianapolis, USA) without size selection. The fragment size structure of the DNA library was assessed using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, USA). The equimolar pool of 4 nM DNA libraries was denatured, diluted to 13 pM and sequenced (200-bp paired-end sequencing) on the Illumina MiSeq platform (Illumina, USA).

High-quality trimmed reads were used for de novo assembly and contigs blasted to the viral genome database (ftp://ftp.ncbi.nih.gov/genomes/Viruses/all_fna.tar.gz) using CLC Genomics Workbench 7.5 and Geneious v.8.1.9. Alternatively, the reads were mapped against the available complete PVM genome sequences.

The 5'-end of the PVM genomes was determined using total RNA as a template for 5' Rapid amplification of cDNA ends experiment (SMARTer® RACE 5'/3' Kit; TaKaRa, Bio Inc., Shiga, Japan) with a gene-specific primer rPVM_509R (5’ GCTCGTAGGACCATAGGGCAC 3’), following the manufacturer’s instructions. The 3'-end of the genome was RT-PCR-amplified using the oligo(dT) primer in combination with an internal specific primer PVM_8367F (5’ GGTATCTTCTTTCGTGCGTCC 3’). In both 5’- and 3’-termini, the overlapping sequences (up to 300 bp) matched unambiguously the HTS-based sequences, confirming their accuracy.

In the case of the divergent isolate T20, Sanger sequencing of PCR products obtained using specific primers designed from the HTS-based sequences was used to close an internal gap in the genome (ca. 2.3 kb between nt 1500 and 3800).

Conventional serological and molecular detection. Leaf samples were obtained during July–September 2016 and 2017 from tomato (Solanum lycopersicum L.) plants growing in private gardens in three geographically distant localities in Slovakia.

Total RNAs were extracted from tomato leaves using the NucleoSpin RNA Plant kit (Macherey-Nagel, Germany). For RT-PCR detection, a two-step protocol was systematically used. The first-strand cDNA was synthesised using random hexamer primers and AMV reverse transcriptase (Promega Corp., Madison, USA). An aliquot of RT reaction was added to PCR carried out with TaKaRa Ex Taq polymerase (TaKaRa, Bio Inc., Japan). A primer pair PVM_d8065F (5’-CTCAATGCACAGGTCACTG-3’) and PVM_d8445R (5’-TGACTGAAGGTATCACCTC-3’), was designed from PVM sequences available in the Genbank database (www.ncbi.nlm.nih.gov; accessed on July 2016) in order to match the actual virus diversity.

The RT-PCR products were gel-purified using the Wizard SV Gel and PCR Clean-Up System (Promega Corp., USA), and directly sequenced in both directions by priming the sequencing reactions with the same oligonucleotides used for PCR amplification.
The presence of PVM and additional viral pathogens in the tomato samples, namely *Potato virus Y* (PVY), *Cucumber mosaic virus* (CMV), *Potato virus S* (PVS), and *Tomato mosaic virus* (ToMV) was checked by DAS-ELISA using a commercially available kit (Bioreba AG, Reinach, Switzerland). In parallel, RT-PCR was performed on the template of total RNAs using protocols for detection of PVY (PVY-9063F/PVY-9316R; Sihelská et al. 2016), PVS (PVS_7833F/PVS_8386R; PREDAJŇA et al. 2017), ToMV (ToMV-F/ToMV-R; Sihelská et al. 2017) and CMV (CMV_CPf 5'-TYTT-CATGGATGCTTCTCCRC-3'/ CMV_CPr 5'-CTGGATGGACAACCGGTTC-3'; Glasa M., unpublished).

**Sequence analyses.** Nucleotide PVM sequences reported in this paper have been deposited in the GenBank database and have the following accession numbers: MH558035–MH558037 (full-length genomes), MH558038–MH558047 (partial genomes).

Sequence analyses and comparisons were done using the MEGA v7 program (Kumar et al. 2016). The alignments were used as input data to construct phylogenetic trees by the neighbour-joining distance method implemented in MEGA; each time the bootstrap analysis with 1 000 replicates was performed to assess the robustness of the branches. DnaSP v.5 (Librado & Rozas 2009) was used to estimate the average intra-group divergence values.

Searches for potential recombination events and identification of recombination breakpoints employed seven methods, including RDP, GENECONV, Bootscan, Maxchi, Chimaera, SiScan and 3SEQ, implemented in the RDP4 v.4.1 software (Martin et al. 2015), using default settings. Recombination events were noted if supported by at least four different methods (*P*-values < 1.0 × 10⁻⁶).

**RESULTS**

Mapping the individual reads (2.1–3.5 millions) obtained from T20, T40, and T50 samples against reference genomes available in the databases (NC_001361) enabled recovery of the nearly complete PVM genome of T40 and T50 isolates (Table 2). In the case of T20,

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Locality</th>
<th>PVM accession number</th>
<th>Viral pathogens identified by HTS</th>
<th>Standard detection (ELISA/RT-PCR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T20</td>
<td>Paňa</td>
<td>MH558035</td>
<td>PVM, PVY</td>
<td>PVM, PVY</td>
</tr>
<tr>
<td>T40</td>
<td>Plavecký Mikuláš</td>
<td>MH558037</td>
<td>PVM, PVY, PVS</td>
<td>PVM, PVY, PVS</td>
</tr>
<tr>
<td>T50</td>
<td>Plavecký Mikuláš</td>
<td>MH558036</td>
<td>PVM, PVY, CMV</td>
<td>PVM, PVY, CMV</td>
</tr>
<tr>
<td>T16</td>
<td>Paňa</td>
<td>MH558038</td>
<td></td>
<td>PVM, PVY</td>
</tr>
<tr>
<td>T18</td>
<td>Paňa</td>
<td>MH558039</td>
<td></td>
<td>PVM, PVY, CMV, PVS</td>
</tr>
<tr>
<td>T32</td>
<td>Plavecký Mikuláš</td>
<td>MH558047</td>
<td></td>
<td>PVM, PVY, CMV</td>
</tr>
<tr>
<td>T48</td>
<td>Plavecký Mikuláš</td>
<td>MH558044</td>
<td></td>
<td>PVM, PVY</td>
</tr>
<tr>
<td>T49</td>
<td>Plavecký Mikuláš</td>
<td>MH558041</td>
<td></td>
<td>PVM, PVY</td>
</tr>
<tr>
<td>T53</td>
<td>Plavecký Mikuláš</td>
<td>MH558043</td>
<td></td>
<td>PVM, PVY</td>
</tr>
<tr>
<td>T57</td>
<td>Plavecký Mikuláš</td>
<td>MH558042</td>
<td></td>
<td>PVM, PVY</td>
</tr>
<tr>
<td>T58</td>
<td>Plavecký Mikuláš</td>
<td>MH558040</td>
<td></td>
<td>PVM, PVY</td>
</tr>
<tr>
<td>T62</td>
<td>Sološnica</td>
<td>MH558046</td>
<td></td>
<td>PVM, CMV, PVS</td>
</tr>
<tr>
<td>T65</td>
<td>Sološnica</td>
<td>MH558045</td>
<td></td>
<td>PVM, PVY</td>
</tr>
</tbody>
</table>

PVM – *Potato virus M*; PVY – *Potato virus Y*; PVS – *Potato virus S*; CMV – *Cucumber mosaic virus*

Table 2. Analysis of the ribosomal-depleted total RNA-derived high-throughput sequencing reads from the tomato samples

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Total reads</th>
<th>PVM mapped readsa</th>
<th>Mapped reads (%)</th>
<th>Average coverage</th>
<th>Reference genome covereda (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T20</td>
<td>3585846</td>
<td>259664</td>
<td>7.2</td>
<td>3866.4 x</td>
<td>73b</td>
</tr>
<tr>
<td>T40</td>
<td>2162142</td>
<td>384465</td>
<td>17.7</td>
<td>5512.1 x</td>
<td>99.8</td>
</tr>
<tr>
<td>T50</td>
<td>2466092</td>
<td>299219</td>
<td>12.1</td>
<td>4198.6 x</td>
<td>99.9</td>
</tr>
</tbody>
</table>

amapped to the PVM NC_001361 genome; bdue to an unmapped gap between nt 1500–3800
the percentage of the reference genome covered by the reads was much lower (73%), due to its sequence divergence respect to the PVM isolate used as reference. In order to recover the full-length sequence of this putative divergent isolate, Sanger sequencing was used to fill a 2.3 kb gap in the open reading frame 1 (ORF1) region of the genome. A subsequent mapping against such completed T20 sequence showed a full-length coverage of almost whole genome (data not shown), confirming a high variability of the T20 isolate in this region of the genome. For all 3 isolates, the 5’ and 3’ extremities were determined by 5’ RACE and oligo d(T)-anchored RT-PCR.

The genomic organisation was fully conserved between all Slovak PVM isolates and was typical of members of the genus *Carlavirus* (Figure 1). The genome was either 8 523 nt (T40 and T50) or 8 524 nt long and contained 6 ORFs. From 5’ to 3’ these correspond to a large replication-associated protein (ORF1, translated to 1964 aa), followed by three ORFs encoding the proteins recognised as triple gene block (TGB, length of 229 aa, 109 aa, and 63 aa, respectively). ORF5 is translated to the coat protein (304 aa) with calculated mass of 33.589, 33.897, and 33.917 kDa for T20, T40, and T50, respectively. The last ORF encoded a nucleic acid binding protein (NABP, 108 aa).

The genome was highly collinear between three isolates, with only one insertion in the T20 genome (7 202 nt) occurring in the intergenic region between ORF4 and ORF5. 5’ non-coding region (NCR) was determined by 5’ RACE and was 76 nt long in all 3 Slovak isolates, while the length of 3’ NCR was 73 nt.

The PVM diversity was analysed using the available dataset of 16 full-length PVM sequences (13 available in the databases and 3 determined in this work). At the whole genome level, the average pairwise divergence among PVM isolates reached 17.3 ± 0.3%. This value is higher in comparison with the known intraspecies variability of a related *carlavirus*, PVS. Computed for a dataset composed of 33 full-length PVS sequences available actually (Sept 2018) in the Genbank database, it shows an 11.4 ± 0.2% divergence. In both viruses, the highest diversity was observed in ORF1 around nt positions 1 500–3 000 (Figure 1).

Figure 2 shows a clear separation of PVM isolates into two major phylogenetic clusters. T40 and T50 isolates, originally collected from unrelated private gardens located in the same village (Plavecký Mikuláš), are phylogenetically related and fell into the major phylogenetic PVM group, together with potato and tomato isolates from Europe and Asia. Isolate YN (KY364848), reported recently as a divergent PVM variant (Su et al. 2017) is included in a well-supported phylogenetic group with Bangladesh isolates (MF133528–MF133530). Interestingly, the Slovak T20 isolate did not cluster together with any of the isolates analysed, either with the major group including T40 and T50 or with the one represented by Chinese and Bangladesh isolates. Thus, T20 represents a distinct lineage in the phylogenetic tree that potentially corresponds to a new phylogenetic group.

The detection of PVM in tomato has prompted further screening of its occurrence on a larger set of 61 samples. Using DAS-ELISA and RT-PCR, 10 out of 61 tomato samples were tested positive for PVM. All of these samples were included in a phylogenetic analysis based on the partial sequences spanning
the 3’ end of the genome (340 bp, 8,086–8,425 nt). This phylogenetic study involving 109 PVM isolates further confirmed the clustering of PVM isolates in a major group and a separate position for PVM T20. With the exception of T20 isolate, all partially sequenced Slovak PVM isolates belong to major Group I, although divided into two subclusters (Figure 3). No evidence of recombination was identified in the dataset of 16 complete PVM genomes using RDP analysis.

The primers used for the molecular detection of the Slovak PVM isolates studied in this work (PVM_d8065F/PVM_d8445R) were designed based on the sequence dataset available in the Genbank (accessed on July 2016), targeting the highly conservative genome portions spanning the 3’ end of the CP gene and 3’ end of the NABP gene, respectively (Figure 4). Despite 3 and 2 mutations, respectively, in the primer sequences, the divergent T20 was amplified by RT-PCR, providing the evidence of polyvalence of these primers for Slovak PVM isolates, and likely for all PVM isolates belonging to Group I.

**DISCUSSION**

The reports of new virus species recovered from tomato plants are increasing (Li et al. 2012; Saqib et al. 2015; Fontenele et al. 2017; Xu et al. 2017; Vaghi Medina et al. 2018). However, attention has also to be paid to cover the real diversity of already established and “common” tomato viruses, which could be still overlooked.

As part of the characterisation of the virome present in cultivated tomato plants in Slovakia, ribosomal-depleted total RNAs from three tomato plants grown in private gardens in the open air (labelled as T20, T40, and T50) were subjected to HTS analysis. Leaf
and nation event was identified in the TGB2-TGB3-CP region of the Chinese PVM isolate from pepino (Ge et al. 2014) and in the CP gene of Blueberry scorch virus (Kalinowska et al. 2015) and PVS (Lin et al. 2014). The lack of the recombination event detected within PVM in this work should thus reflect a still insufficient number of complete genome sequences available, not representative of all genetic variants identified previously by partial sequencing.

Conventional detection techniques based on RT-PCR have increased the sensitivity and accuracy of viral diagnostics, however, they only allow the detection of known viruses, and their development and need of optimisation depend on the availability of sequence data. The current knowledge and complexity of molecular data for a given virus can still be biased or incomplete not only in recently identified virus pathogens, but also in common and long-time established viruses. Consequently, developed detection techniques need not polyvalent enough, resulting in the false negative diagnostic results for divergent variants or strains.

As shown before, recently reported PVM isolates from China and Bangladesh represent a distinct genetic group (Figures 2 and 3). In silico analysis of their sequences revealed several mutations in the binding site of the forward primer (Figure 4). As two of these mutations affect positions close to the primer 3’ end, the suitability of this primer for detection of Group II isolates might be negatively affected. Therefore, a reliable RT-PCR-based detection of Group II PVM isolates should be developed, however, probably only when more sequence data are available.

PVM is considered mostly latent or causing only moderate symptoms in infected potatoes (Brunt et

samples were collected during the vegetation period from plants showing either virus-like symptoms (mild leaf distortions in T20 and leaf deformations and growth reduction in T50) or having a symptomless behaviour (T40).

Blast analysis of de novo assembled contigs revealed the presence of PVM in all 3 tomato plants, in a complex infection involving PVY, CMV and/or PVS (Table 1). To our knowledge this is the first report and characterisation of PVM in tomato in Slovakia.

The relatively high intraspecies variability detected for PVM (17.3%) is mainly due to the recent characterisation of a divergent Chinese isolate YN (KY364848; Su et al. 2017) and isolates from Bangladesh (MF133528–MF133530; Xu H., direct submission). Interestingly, these divergent PVM isolates were characterised using HTS, further confirming the potential of this approach for characterisation of isolates escaping standard detection methods. In this line, the identification of Slovak T20 isolate and determination of its complete genome sequence would be highly challenging using standard detection and sequencing methods, due to its molecular distinctness from other Slovak PVM isolates.

Recombination represents one of the key evolutionary processes underlying the high genetic variability of plant RNA viruses, contributing to the macro-evolution of viruses and to the extent of their intraspecific diversity (Sztuba-Solińska et al. 2011). Within the genus Carlavirus, recombination was reported to occur in the evolutionary history of Cowpea mild mottle virus (Zanardo et al. 2014) and Chrysanthemum virus B (Singh et al. 2012). Similarly, using partial sequence data, a recombination event was identified in the TGB2-TGB3-CP region of the Chinese PVM isolate from pepino (Ge et al. 2014) and in the CP gene of Blueberry scorch virus (Kalinowska et al. 2015) and PVS (Lin et al. 2014). The lack of the recombination event detected within PVM in this work should thus reflect a still insufficient number of complete genome sequences available, not representative of all genetic variants identified previously by partial sequencing.

Conventional detection techniques based on RT-PCR have increased the sensitivity and accuracy of viral diagnostics, however, they only allow the detection of known viruses, and their development and need of optimisation depend on the availability of sequence data. The current knowledge and complexity of molecular data for a given virus can still be biased or incomplete not only in recently identified virus pathogens, but also in common and long-time established viruses. Consequently, developed detection techniques need not polyvalent enough, resulting in the false negative diagnostic results for divergent variants or strains.

As shown before, recently reported PVM isolates from China and Bangladesh represent a distinct genetic group (Figures 2 and 3). In silico analysis of their sequences revealed several mutations in the binding site of the forward primer (Figure 4). As two of these mutations affect positions close to the primer 3’ end, the suitability of this primer for detection of Group II isolates might be negatively affected. Therefore, a reliable RT-PCR-based detection of Group II PVM isolates should be developed, however, probably only when more sequence data are available.

PVM is considered mostly latent or causing only moderate symptoms in infected potatoes (Brunt et
al. 2001), however, less data are available for PVM infection in tomato. Although some of the tomato plants tested in this work did not show clear virus-like symptoms, several of them exhibited leaf mosaics, yellowing, and apical leaf deformation (data not shown). As the PVM-positive tomato plants were found to be also infected with CMV, PVS or PVY in double or triple infection (Table 1), it is not possible, however, to assess the potential contribution of PVM to the symptoms observed.

PVM is one of the “common” pathogens that have been recognised to infect plants for a long time (Bagnall & Larson 1957; Brunt et al. 2001). Despite of this fact, the knowledge of its variability was limited, mainly due to the reduced availability of full-length genomes.

In this work, three Slovak PVM isolates infecting tomato have been sequenced and characterised. A phylogenetic analysis of these three full-length sequences together with other sequences available in the databases has revealed the existence of a new phylogenetic divergent group of this virus. Our results further pinpoint the significance of high-throughput sequencing approaches providing an unbiased access to viral diversity studies, with the potential consequences for the development/optimisation of routine pathogen detection.

References

Bagnall R.H., Larson L.H. (1957): Potato virus M. Phytopathology, 47: 3


