Molecular variability in the cysteine rich protein of potato virus M

Fatemeh Tabasinejad¹ · Behrooz Jafarpour¹ · Mohammad Zakiaghl¹ ·
Majid Siampour² · Hamid Rowhani¹ · Mohsen Mehrvar¹

Received: 9 March 2015 / Accepted: 10 June 2015 / Published online: 26 June 2015
© Indian Virological Society 2015

Abstract The potato virus M (PVM), belonging to the genus *Carlavirus*, is a worldwide endemic pathogen in potato fields. p11 is an 11–16 kDa protein encoded by the last open reading frame of PVM which contains cysteine rich proteins (CRPs) motif. CRPs have been identified as suppressors of gene silencing. In this study the p11 gene from 28 PVM isolates, including 16 new isolates from Iran, were used to determine the global genetic structure of PVM populations. Pairwise nucleotide sequence identity scores showed that global PVM CRP sequence similarity was between 69.3 and 100 %. This genetic diversity divided the 28 isolates into two main divergent phylogenetic clades. The rate of genetic diversity and non-synonymous to synonymous mutations (dN/dS) were significantly different between these two clades. Analysis showed that PVM CP is under significant negative selection pressure with the global ω value of 0.260.

Keywords Potato virus M · Cysteine rich proteins · Genetic diversity · Negative selection · Iran

Introduction

The study of the variability and the genetic structure in plant virus populations is an important topic in plant pathology toward understanding of their ecology and epidemiology and has potential to develop strategies for control of viral diseases [3]. On the other hand, characterization of the population genetic structure may provide information to understand evolutionary forces shaping virus populations.

Analysis of viral genome revealed that each gene in the genome is imposed by different selection pressures. The fate of a new variant in virus population depends on different selection pressures concerned on its evolution, hence it is important to understand the evolutionary factors involved in emergence of new genotype in virus populations [5].

Cysteine-rich proteins (CRPs) encoded by plant viruses in the families' *Virgaviridae* and *Flexiviridae* are believed to act as the RNA silencing suppressors [9]. CRP proteins contain a cluster of amino acids similar to the zinc-finger motif, suggesting RNA and/or DNA-binding activity for them [8]. Based on structural and functional similarities; the CRPs can be divided into two groups. First group includes CRPs of the genera *Hordeivirus*, *Tobravirus*, *Benyvirus*, *Pecluvirus*, and *Furovirus* [8]. A characteristic feature of this group is that their CRPs are involved in regulating viral replication, modulate systemic trafficking and act as pathogenicity determinants [17]. The CRPs of *Allexivirus*, *Vitivirus* and *Carlavirus* fall into the second group. These CRPs have a conserved arginine-rich domain and a zinc-finger motif, involved in viral RNA accumulation, cell-to-cell movement and suppression of RNA silencing [8].

Potato virus M (PVM) belongs to the genus *Carlavirus*, encode an 11 kDa protein (p11) containing cysteine-rich motif at the 3' proximal ends [23]. CRP protein of the PVM have RNA and DNA-binding activities in vitro [8], also, it has

Electronic supplementary material The online version of this article (doi:10.1007/s13337-015-0258-8) contains supplementary material, which is available to authorized users.

✉ Mohammad Zakiaghl
zakiaghl@ferdowsi.um.ac.ir

¹ Department of Plant Pathology, College of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran

² Department of Plant Protection, College of Agriculture, Shahrekord University, Shahrekord, Iran

been recently shown that CRP and triple gene block protein 1 of PVM have local and systemic silencing suppression activities respectively [17]. Although p11 protein plays an important role in pathogenicity of PVM, little information is available for diversity of this protein in the PVM population.

Previously, we showed that all PVM isolates evolve in two distinct divergent groups based on analysis of their coat protein gene [18]. The objective of this work is to investigate the diversity of the CRP protein in the PVM population and to study force acting on evolution of this gene in the PVM population.

Materials and methods

Collection and serological diagnosis of virus

During the growing season of 2011, 500 leaf samples were collected randomly from main potato fields in Hamadan, Northern khorasan, Khorasan Razavi, Esfahan and Kerman provinces of Iran. Fresh leaf extract of each sample was subjected to double-antibody sandwich enzyme linked immune sorbent assays (DAS-ELISA) [1] using polyclonal antiserum of PVM. Antiserum and the positive control of PVM were provided from DSMZ (Germany).

RNA extraction and RT-PCR

Total RNA was extracted from the ELISA positive samples (50 mg/sample) using SV total RNA isolation kit according to manufacturer's instruction (Promega, USA). cDNA was synthesized using RevertAidTM M-MuLV reverse transcriptase (Fermentas, Lithuania) and PVM-UTR as reverse primer. A pair of specific primers, PVM-ORF6-F (5'-ATGAAGCACGTAACCAAGGTG-3') and PVM-UTR (5'-TGGTTTTTA ACTATTTT TAGCCAA-3') were designed to amplify a 398 bp (includes complete length of ORF6 (326 bp) and 72 bp of 3'-UTR of PVM genome corresponding to nucleotides 8126–8524) using Emerald Amp[®] MAX HS PCR Master Mix (Takara) following the manufacturer's protocol. The reaction mixture had a final volume of 25 µl and contained 10 pmol of forward and reverse primers and four µl of cDNA extract. The PCR thermal profile were initial denaturation at 95 °C for 4 min, followed by 30 cycles of 94 °C for 30 s, 51 °C for 45 s and 72 °C for 30 s, with a final extension at 72 °C for 5 min. The amplified DNA fragments were visualized in 1 % agarose gel containing 10 µg/ml DNA green-viewer.

Molecular cloning and sequencing

The PCR products were purified using the Qiaquick gel extraction kit (Qiagen) and ligated into pTZ57R/T

according to the instructions of the supplier (Fermentas). Competent cells of *Escherichia coli* strain DH5α were transformed by recombinant plasmids and transformants were selected on LB-Agar plates containing ampicillin, IPTG and X-gal. Bacterial cells harboring the recombinant plasmids were grown at 37 °C.

The recombinant plasmids were purified using Prime Prep Plasmid DNA Isolation Kit (Genetbio, Korea) and sequenced in both direction using M13 universal primers by an ABI PRISM system with Macrogen (South Korea). The resulting sequences were compared with those deposited in the GenBank using BLAST program.

Phylogenetic analysis

Complete nucleotide sequences of the p11 gene related to 28 PVM isolates, including 16 new Iranian and 12 PVM isolates from other countries (Supplementary Table 1) were used for phylogenetic analysis. Nucleotide sequences were aligned using muscle implemented in molecular evolutionary genetics analysis package (MEGA, version 6.06). Phylogenetic analysis performed using maximum likelihood method in MEGA 6.06 using GTR+G3 and Jones–Taylor–Thornton (JTT) as the best-fitted models at the nucleotides and the amino acids level, respectively. Statistical confidence and robustness of the node of the phylogenetic trees was assessed by 1000 bootstrap replicates and nodes with bootstrap values <50 % were condensed [20].

The sequences were analyzed in pairwise comparisons to construct similarity color-based identity matrix of the nucleotide and deduced amino acid sequences using SDT v1 software [13]. The mean distance values for PVM ORF6 were computed using the Tamura–Nei model in MEGA V. 6.06 [20]. Number of segregation sites and amount of π (the average number of nucleotide differences between two random sequences in a population) were estimated using DnaSP version 5.10 [7].

To understand the evolutionary constraints imposed on this part of PVM genome, the ratio of dN/dS (the average number of non-synonymous substitutions [dN] per synonymous substitutions [dS]) was estimated by DnaSP version 5.10.

Results

Genetic variability in PVM populations

PVM infection was confirmed, using DAS-ELISA, in 30 out of 500 potato samples. Among five provinces of Iran, Kerman was found as the most infected region (28/180) whereas Fars had no positive samples (0/30).

Using PVM specific primers in RT-PCR, a 398 bp DNA fragment was amplified from the ELISA positive samples (data not shown). Analysis of sequencing data revealed that the entire fragment consisted of 326 nucleotide coding for PVM-ORF6 followed by 72 nt of the 3'UTR.

Pairwise nucleotide sequence comparisons of p11 gene (ORF6) from 16 Iranian PVM isolates with all available in the GenBank ($n = 28$) (Supplementary Table 1) revealed a global diversity between 71.2 and 100 % (Supplementary Fig. 1).

The p11 sequence distance among all PVM isolates showed maximum divergence of 28.7 at nucleotide level and 30.6 % at amino acid level with an average of 13.3 ± 3.96 % and 13.52 ± 4.04 %, respectively (Supplementary Fig. 1). The average divergence among Iranian isolates was 12.64 ± 5.78 % and 12.9 ± 5.74 % with the maximum divergence of 25.1 and 27.8 % at the nucleotide and amino acid levels, respectively (Supplementary Fig. 1). Among the Iranian isolates, the most divergent isolate appeared to be isolate 490 [KC699721]. p11 nucleotide sequence identity scores between Iranian and PVM isolates retrieved from GenBank ranged between 72.2 % (between isolates 691 [KC699727] and Japan [AB608989]) and 99.7 % (between isolates 352 [KC699715] and DSMZ-PV0273 [EU604672]). These data suggest the p11 diversity of the Iranian isolates is comparable with the global diversity (71.2–100 %).

The amino acids identity level of the p11 protein of 16 Iranian isolates ranged between 73.1 and 99.1 % while the global diversity ranged from 71.3 to 100% (Supplementary Fig. 1). Alignment of the p11 amino acid sequences of all PVM isolates revealed that most of the variations along the p11 concentrated in their N-terminal, particularly in a block of 18 amino acids between positions 11 and 29. By contrast, amino acids at the C-terminal were mainly conserved (data not shown).

Phylogenetic analysis of PVM isolates

Nucleotide and amino acid sequence alignments of p11 gene from PVM isolates was used to generate phylogenetic trees by maximum likelihood method. The topology of both trees generated using amino acid and nucleotide sequences of p11 was identical. Phylogenetic analysis clearly indicated that all PVM isolates fall into two main distinct subpopulations, I and II, supported by high bootstrap values (Fig. 1). These two major clades consisted of 20 and 8 PVM isolates, respectively. As shown in the phylogenetic tree, the majority of PVM isolates are arranged within the two groups irrespective of their country of origin. Of 16 isolates collected from Iran, 10 isolates were clustered with members of subpopulation-I and six were clustered with the members of subpopulation-II (Fig. 1). This suggests that geographical origin is not the only determinant factor in the early phylogenetic divergence of PVM isolates into two main clades.

The sequence identity between members of subpopulations I and II ranged from 71.3 to 81 % at the nucleotide and from 69.4 to 78.7 at the amino acid levels. Isolates within the same group (either I or II) shared over 90.2 and 88 % identity at the nucleotide and amino acid level respectively (Fig. 1; Supplementary Fig. 1).

Analysis of the frequency of pairwise sequence identity scores (378 pairwise scores for 28 isolates) for p11 gene sequences at 1 % intervals identified two distinct frequency distributions, one at 96 % for subpopulation-I and the other at 77 % for subpopulation-II. Within each group, the pairwise frequency distribution concentrated at 95 %, thus we used sequence identity at a greater level (96.2 %) for subclade identification. This, coupled with the phylogenetic tree, identified nine subclades in subpopulation I and three subclades for subpopulation II in PVM population (Fig. 1). Isolates from Asia and Canada were found in both subpopulations I and II, but those from Europe were found in four subclades in the subpopulation-I. Six Iranian isolates formed two divergent subclades in subpopulation-II whereas isolates of Canada and Japan deposited in the separated subclades (Fig. 1).

Evolutionary constraints imposed on PVM population

The population statistics were determined and compared among p11 genes of the PVM isolates (Table 1). In total, 125 variable sites were found across all isolates in the nucleotide sequence alignment. The frequency of sequence variation within subpopulation I and subpopulation II were 0.034 and 0.134 respectively. Nucleotide diversity was 3.9 times higher in subpopulation II than in subpopulation I. The population mutation rate, θ in subpopulation II of PVM was 2.2 times higher than that in subpopulation I.

To determine the evolutionary forces acting on the PVM p11 gene, the ω value (dN/dS) was calculated. The dN/dS ratios for the p11 genes were found to be less than 1.0, indicating that they were subjected to negative or purifying selection (Table 1). The global ω value for all of the PVM p11 gene sequences was 0.260. Pairwise comparisons revealed that these values were also significantly less than 1 for most of the CRP sequence pairs (data not shown). The global dN/dS values for the subpopulation I and subpopulation II of PVM was compared. The p11 genes of isolates in the subpopulation I and subpopulation II of PVM had dN/dS values of 0.263 and 0.160 respectively. Statistical analysis showed the selection pressure acting on the p11 genes of the subpopulation I and subpopulation II of PVM is significantly different. In this regard, the dN/dS ratios for the subpopulation II was 1.64 times higher than subpopulation I (Table 1), suggesting that the proteins encoded by subpopulation II are considerably more tolerant to amino acid substitutions.

Fig. 1 Phylogenetic analysis at the amino acids level for CRP protein of 28 PVM isolates. The maximum likelihood tree was constructed using MEGA 6.06 with Jones–Taylor–Thornton (JTT) model as the best fitted model with 1000 bootstrap replicates and the pairwise deletion. Viruses names, acronyms and GenBank accession numbers are in Table 1

Table 1 Summary of population genetics parameters of the CRP genes within PVM population

	m	S	Hd	π	$\theta (S)$	nS	dS	nN	dN	ω	Rm
Total	28	125	0.968 (0.024)	0.124 (0.0191)	0.130	76.11	0.3237	226.88	0.0841	0.260	24
Group 1	21	57	0.947 (0.044)	0.034 (0.0045)	0.056	76.12	0.0812	226.87	0.0213	0.263	8
Group 2	8	87	0.929 (0.084)	0.134 (0.0187)	0.125	81.87	0.4474	242.12	0.0716	0.160	11
worldwide isolates	12	107	1.000 (0.034)	0.1025 (0.0361)	0.141	76.09	0.2570	226.90	0.0691	0.269	8
Iranian isolates	16	104	0.900 (0.062)	0.1332 (0.0208)	0.119	81.72	0.3430	242.27	0.0924	0.270	15
European isolates	8	33	1.000 (0.63)	0.0320 (0.0053)	0.042	76.14	0.0824	226.85	0.0172	0.210	2
Asian isolates	3	90	1.000 (0.272)	0.1845 (0.0841)	0.249	81.05	0.5335	242.94	0.1286	0.241	0

m number of sites, *S* number of segregating sites, *Hd* measurement of haplotype diversity and its variance, π nucleotide diversity, $\theta (S)$ statistic θ from segregating site, *nS* number of synonymous sites, *dS* Synonymous substitution, *nN* number of nonsynonymous sites, *dN* nonsynonymous substitution, ω ratio of dN/dS, *Rm* minimum number of recombination events

Discussion

To date, Potato virus S (PVS), Potato virus H (PVH), Potato latent virus (PotLV), Potato virus P (PVP) and PVM are known carlaviruses reported to infect potato plants naturally worldwide [4, 6, 10, 14].

The objective of this research was to characterize the genetic diversity of p11 (CRP) protein in the PVM

populations. In the present study, we examined the genetic diversity of 28 PVM isolates originated from seven countries (including Canada, China, Czech Republic, Poland, Russia, Germany, Japan and Iran) by targeting their p11 gene sequences.

Sequence comparisons of the p11 gene revealed considerable similarity at the nucleotide and the amino acid levels ranging from 69.3 to 100 % among PVM isolates.

Similar results have been reported for other carlaviruses including of PVS, PVM, PotLV, PVP and PVH based on the coat protein gene [4, 6, 10, 14].

Maximum likelihood based phylogenetic trees constructed using nucleotide or amino acid sequences of the p11 gene resembled identical topology. Phylogenetic analysis of the p11 gene identified two putative divergent groups of isolates supported by high bootstrap values. Each of the putative groups included isolates from Iran clustered with isolates from other countries. Thus, PVM isolate variation does not appear to be associated with the geographic origin. The lack of association of isolate variation with geographic location highlights the world-wide movement of potato germplasm. Similar results have been reported for PVS [2, 16], PVM [18, 21], PVP [6] and *Hop mosaic virus* (HpMV) [15].

The second objective of this study was to determine the evolutionary forces acting on PVM populations. We determined that mutation and purifying selection were the predominant evolutionary forces influencing the population structure of PVM population.

Haplotype analysis is an indicator to measure the diversity between sequences [3]. This parameter is helpful to assess the genetic diversity in viral populations. The measures of haplotype diversity, Hd, may range from 0 (no diversity) to 1 which specifies high levels of haplotype diversity [12].

Haplotype diversity of the p11 gene in PVM subpopulations were 0.900–1.000, indicating very high levels of diversity in p11 gene for all subpopulations. In addition, Hd were 0.947 and 0.929 for two subpopulations of CRP genes in PVM population. These levels of diversity were comparable to that estimated for global populations (0.968), supporting that the haplotype diversity may not affect with geographic location.

Also, we used s , the number of segregating sites, Waterson's estimate of θ and the average number of nucleotide differences π , to state of the subpopulation [19]. Surprisingly, these parameters were different within two subpopulations of PVM. A fourfold difference was observed in the average number of nucleotide differences π , between subpopulations. The selection pressure on p11 gene, dN/dS or ω value, was calculated to compare synonymous to nonsynonymous nucleotide substitution rates. The ω for the p11 gene was significantly <1 in the PVM subpopulations, which indicated strong purifying selection on the p11 gene of PVM [22].

Analysis also showed that the CRP protein in the subpopulation-II is under greater selection pressure (Up to 1.64 times) than CRP in subpopulation I, indicating that the ratio of nonsynonymous to synonymous mutations in the CRP gene is higher in the earlier. Despite higher ω value, the level of CRP sequence identity among isolates in

subpopulation-I was considerably higher than that in subpopulation-II.

Phylogenetic and population differentiation analysis based on the CRP gene of PVM population at the nucleotide and amino acids levels revealed separation of PVM isolates into two divergent clusters. Previously, we showed different isolates of PVM have been segregated into two divergent evolutionary lineages of PVM-o (PVM-ordinary) and PVM-d (PVM-divergent) based on analysis of the CP gene, [18]. Surprisingly, this differentiation was also largely confirmed by analysis of the PVM CRP gene.

Population dynamics of carlaviruses have been poorly understood. Inspection in populations of PVS, closely related to the PVM, has differentiated its isolates into two biologically different groups, which is in accordance with phylogenetic analysis of its CP gene [2, 11, 16]. Populations of HpMV, another *Carlavirus*, were clustered into two phylogenetically distinct groups based on their nucleic-acid-binding proteins sequence, although these isolates were placed into three divergent groups based on the HpMV CP gene [15].

We believe that mutation and geographical isolation are factors or mechanisms that build up the genetic diversity of PVM populations. Negative selection turns in the reverse direction to maintain amino acid conservation of the CRP gene in the population. Analysis of additional PVM genes is recommended for better understanding of the genetic structure and variation of PVM populations on a large geographical scale.

References

1. Clark MF, Adams A. Characteristics of the microplate method of enzyme-linked immunosorbent assay for the detection of plant viruses. *J Gen Virol*. 1977;34:475–83.
2. Cox BA, Jones RAC. Genetic variability in the coat protein gene of Potato virus S isolates and distinguishing its biologically distinct strains. *Arch Virol*. 2010;155:1163–9.
3. García-Arenal F, Fraile A, Malpica JM. Variability and genetic structure of plant virus populations. *Annu Rev Phytopathol*. 2001;39:157.
4. Goth RW, Ellis PJ, de Villiers G, Goins EW, Wright NS. Characteristics and distribution of Potato Latent Carlavirus (Red LaSoda Virus) in North America. *Plant Dis*. 1999;83:751–3.
5. Li H, Roossinck MJ. Genetic bottlenecks reduce population variation in an experimental RNA virus population. *J Virol*. 2004;78:10582–7.
6. Li YY, Zhang RN, Xiang HY, Abouelnasr H, Li DW, Yu JL, McBeath JH, Han CG. Discovery and characterization of a novel Carlavirus infecting potatoes in China. *PLoS One*. 2013; 8:e69255.
7. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*. 2009;25:1451–2.
8. Lukhovitskaya N, Soloviyev A, Koshkina T, Zavriev S, Morozov SY. Interaction of the Carlavirus cysteine-rich protein with the plant defense system. *Mol Biol*. 2005;39:785–91.

9. Martelli GP, Adams MJ, Kreuze JF, Dolja VV. Family Flexiviridae: a case study in virion and genome plasticity. *Annu Rev Phytopathol.* 2007;45:73–100.
10. Massa GA, Segretin ME, Colavita M, Riero MF, Bravo-Almonacid F, Feingold S. Biological and sequence data suggest that potato rough dwarf virus (PRDV) and potato virus P (PVP) are strains of the same species. *Arch Virol.* 2006;151:1243–7.
11. Matousek J, Schubert J, Dedic P, Ptáček J. A broad variability of potato virus S (PVS) revealed by analysis of virus sequences amplified by reverse transcriptase polymerase chain reaction. *Can. J Plant Pathol.* 2000;22:29–37.
12. Moury B, Desbiez C, Jacquemond M, Lecoq H. Genetic diversity of plant virus populations: towards hypothesis testing in molecular epidemiology. *Virus Res.* 2006;67:49–87.
13. Muhire B, Martin DP, Brown JK, Navas-Castillo J, Moriones E, Zerbini FM, Rivera-Bustamante R, Malathi VG, Briddon RW, Varsani A. A genome-wide pairwise-identity-based proposal for the classification of viruses in the genus Mastrevirus (family Geminiviridae). *Arch Virol.* 2013;158:1411–24.
14. Nisbet C, Butzonitch I, Colavita M, Daniels J, Martin J, Burns R, George E, Akhond MAY, Mulholland V, Jeffries CJ. Characterization of Potato rough dwarf virus and Potato virus P: distinct strains of the same viral species in the genus Carlavirus. *Plant Pathol.* 2006;55:803–12.
15. Poke FS, Crowle DR, Whittock SP, Wilson CR. Molecular variation of hop mosaic virus isolates. *Arch Virol.* 2010;155:1721–4.
16. Salari Kh, Massumi H, Heydarnejad J, Hosseini pour A, Varsani A. Analysis of Iranian Potato virus S isolates. *Virus Genes.* 2011;43:281–8.
17. Senshu H, Yamaji Y, Minato N, Shiraishi T, Maejima K, Hashimoto M. A dual strategy for the suppression of host antiviral silencing: two distinct suppressors for viral replication and viral movement encoded by Potato virus M. *J Virol.* 2011;85:10269–78.
18. Tabasinejad F, Jafarpour B, Zakiaghil M, Siampour M, Rowhani H, Mehrvar M. Genetic structure and molecular variability in Potato virus M populations. *Arch Virol.* 2014;159:2081–90.
19. Tajima F. The effect of change in population size on DNA polymorphism. *Genetics.* 1989;123:597–601.
20. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0.6. *Mol Biol Evol.* 2013;30:2725–9.
21. Xu H, D'Aubin J, Nie J. Genomic variability in Potato virus M and the development of RT-PCR and RFLP procedures for the detection of this virus in seed potatoes. *Virol J.* 2010;7:25.
22. Yang Z, Nielsen R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol.* 2000;17:32–43.
23. Zavriev S, Kanyuka K, Levay K. The genome organization of Potato virus M RNA. *J Gen Virol.* 1991;72:9–14.