Original Article   Open Access

Evolutionary features of 8K (KDa) silencing suppressor protein of Potato mop-top virus

Aminallah Tahmasebi[1,*], Alireza Afsharifar[1], Ahmad Heydari[2], Mohammad Mehrabadi[3]

1) Plant Virology Research Center, College of Agriculture, Shiraz University, Shiraz, Iran
2) Firoozabad Center for Higher Education, Iran
3) Department of Entomology, Faculty of Agriculture, Tarbiat Modares University, Tehran, Iran

ABSTRACT

The cysteine-rich 8K protein of Potato mop-top virus (PMTV) suppresses host RNA silencing. In this study, evolutionary analysis of 8K sequences of PMTV isolates was studied on the basis of nucleotide and amino acid sequences. Twenty-one positively selected sites were identified in 8K coding regions. Recombination events were found in the 8K of PMTV isolates with a rate of 1.8. Totally 30 haplotypes were detected with haplotype diversity ranging from 0.8 to 1.0 and nucleotide diversity from 7.58 to 13.62. The positions 33 and 30 indicated the highly positive and negative selection (with the highest and the lowest dN-dS values), respectively. Tajima's D test suggested that 8K is evolving with a strong positive selection for worldwide isolates. High frequency of segregating sites was identified along 204 positions of *8K*. Moreover, in this study, we used Shannon entropy-based approach to evaluate the variability of each site of nucleotide and corresponding amino acid. Based on Shannon entropy method, 139 and 97 nucleotide sites had the highest entropy value, while 47 and 33 amino acid sites showed the most diversity along 8K sequences. Our findings suggest that 8K as an RNA silencing suppressor evolves rapidly. Taken together, its variability might play a big threat to infect other plants or overcome resistant cultivars.

Keywords: Genetic variation; 8K; Potato mop-top virus; RNA silencing suppressor

INTRODUCTION

Potato mop-top virus (PMTV) belongs to the *Pomovirus* genus in the *Virgaviridae* family [1]. PMTV occurs in many potato-production areas of the world. PMTV has been detected from many parts of the world including Americas, parts of Asia and Europe [2-4]. The particles of PMTV are tubular rod-shaped with a tripartite, positive-sense and single-stranded RNA genome [5]. The tripartite genome of PMTV contains eight open reading frames (ORFs). RNA1 and RNA2 encode virus replication and encapsidation/transmission functions, respectively [6, 7]. A triple gene block (TGB) and 8K are encoded by RNA3 [5]. TGB-encoded proteins cause the cell-to-cell movement of PMTV [8, 9]. The fourth protein, 8K increases the virulence of the virus and acts as an RNA silencing suppressor [10].

*Corresponding Author: Plant Virology Research Center, College of Agriculture, Shiraz University, Shiraz, Iran
Tel: +98-71-32286154
E. mail: atahmasebi@shirazu.ac.ir

PMTV causes brown arcs and circles in potato tubers, thereby it is considered as a big threat to make the potato tubers unmarketable [2, 3]. PMTV isolates from Europe, North America and Asia share a high level of sequence identity [11-16]. Viral pathogenicity has often been related to viral suppressors of RNA silencing (VSR) [17, 18]. VSR proteins have been evolved by viruses to counter RNA silencing based-defense. RNA silencing acts as a potent antiviral defense mechanism to protect plants against viruses [19]. It was shown that the relative rate of VSRs evolution is faster than other virus genes [20]. PMTV *RdRp*, *CP*, *CP-RT* and *TGB* genes showed negative selection, while the *8K* was under positive selection. Also, the *8K* sequence of the PMTV isolates indicated the lowest identity among PMTV genes [21]. However, there is little information about evolutionary features of 8K as a VSR in worldwide isolates. Therefore, the evolutionary analysis of 8K suppressor is essential to develop better strategies for controlling PMTV infection.

In the present study, the complexity of 8K was characterized using two indices: the mean nucleotide diversity and the Shannon entropy. Shannon entropy was originally developed by Claude Shannon and it is also used to measure the information content or complexity of a series. Shannon information is regularly used to measure species diversity and population genetics [22, 23]. Also, the mean nucleotide diversity measures the average number of nucleotide differences between any two sequences [24].

MATERIALS AND METHODS

Multiple-sequence alignment: Multiple-sequence alignment was performed among worldwide isolates using CLC Main Workbench 7.7.2 software. A number of 72 sequences of 8K were then aligned, concatenating triplets according to the sequence alignment using the MEGA program, version 6.06. The model of nucleotide or amino acid substitutions was evaluated by a maximum likelihood approach. Substitution pattern and rates were estimated with the different models.

Non-synonymous (dN) and synonymous (dS) substitution rates are estimated using the joint Maximum Likelihood reconstructions under a Muse-Gaut model of codon substitution and Felsenstein model of nucleotide substitution. Mean (relative) evolutionary rate was estimated using the Jones-Taylor-Thornton model (+G). A discrete Gamma (+G) distribution was applied to model evolutionary rate differences among sites. The substitution patterns (transitional and transversional substitution rates) and nucleotide frequencies for 8K gene were estimated based on the Tamura-Nei model.

Comparative analysis of 8K sequence evolution: To identify specific amino acid sites under selective constraints, the number of differences between dN and dS substitution rates were estimated for each position in the alignments using the MEGA program, version 6.06. A dN-dS value of > 0 is considered as evidence for positive selection, while values of < 0 are indicative of negative selection. The statistical tests for genetic differentiation, estimation of a number of haplotypes and haplotype diversity was performed using DnaSP 5.10.01 software [25]. Potential recombination events between nucleotide sequences of 8K sequences were assayed using DnaSP 5.10.01 software.

Entropy analysis: Entropy analysis was used to evaluate the variability and complexity of each site of nucleotide and amino acid. Shannon entropy was calculated using the definition [23]: $H_i = - \sum_j p_{ij} \log_2 p_{ij}$, *for each site i*, where p_{ij} is the frequency of nucleotide or amino acid j in the site i and H_i is the entropy of each site i. Index j is equal to 1,2,3 or 4 corresponding to nucleotides A, C, G and T respectively or an integer among 1,2,…,20 corresponding to different codons of amino acids alphabet. Sites with entropy value equal to 0 are defined as those which are without complexity and diversity.

RESULTS

The substitution pattern for *8K* gene based on the Tamura-Nei model showed that the transitional substitution rates were from 8.69 to 28.06, while transversional substitution rates varied from 0.93 to 3.02. Compared to transversional substitutions, high transitional substitutions were identified along the *8K* region. The nucleotide frequencies of *8K* gene were: 21.61% (A), 41.58% (T/U), 12.87% (C), and 23.94% (G). The ratios of transition and transversion were k1= 14.747 (purines) and k2= 9.301 (pyrimidines). The transition/transversion bias was R= 5.083 and R= [A*G*k1 +T*C*k2]/[(A+G) *(T+C)]. Mean evolutionary rates were indicated for each 8K amino acid position (Fig. 1).
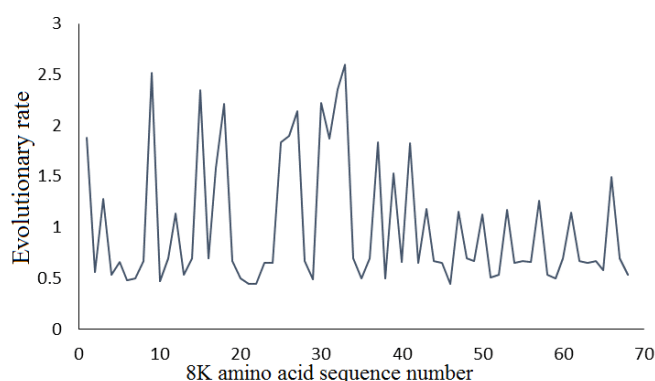

Figure 1: Mean (relative) evolutionary rates over 8K amino acid sequence. Mean (relative) evolutionary rates are shown for each site next to the site number. The average of evolutionary rate across all sites is 1. This means that sites showing a rate <1 are evolving more slowly than average whereas those with a rate>1 are evolving faster than average. The analysis involved 72 amino acid sequences.

The average evolutionary rate across all positions is considered 1; a rate < 1 shows that site evolves slower than the average while the sites with a rate> 1 evolve faster than the average. These rates were estimated using the Jones-Taylor-Thornton model (+G) [26]. The differences in evolutionary rate among sites were based on a discrete Gamma (+G) distribution model. The ML estimate of the gamma parameter was 0.7798. The maximum Log likelihood for this computation was -464.203. Recombination events were identified in the 8K of PMTV isolates with average value; 1.8. Seven recombination sites were detected between nucleotides of (1,25); (25,43); (74,80); (80,89); (89,90); (95,121); (121,139). The degree of negative selection of genes can be estimated by Tajima's D and dN-dS approaches. Tajima's D method computes the differences between the number of segregating sites and neutral expectation. Tajima's D test suggested that 8K is evolving with a strong positive selection for worldwide isolates. The dN-dS approach evaluated the selection by differences between nonsynonymous and synonymous positions (dN-dS) of *8K* sequences. The dN-dS value below zero is representing negative selection. Whereas, a dN-dS value above zero is considered as positive selection. In this study, the mean of dN-dS values showed heterogeneous dN-dS values (from -3.29 to 1.82). The positions 33 and 30 indicated the highly positive and negative selection (with the highest and the lowest dN-dS values), respectively (Fig. 2).

Totally, a number of 21 positively selected sites were identified using site-by-site analysis of the dN/dS ratio. Also, 30 haplotypes were detected and haplotype diversity varied from 0.8 to 1.0 and nucleotide diversity from 7.58 to 13.62. High frequency of segregating sites (42 segregating sites and more than 20.59 % of the total number of sequence bases) were identified along 204 positions of *8K* (Table 1). Alignment of 8K sequences identified some amino acid substitutions among them. A change was found in the start codon of 8K in some isolates. The replacement of the AUG start codon by the GUG occurred. 9.72 percent of the isolates uses

AUG (methionine amino acid) as a translation initiation codon, while GUG (valine amino acid) is replaced in other isolates (90.28 %) (Fig. 3).
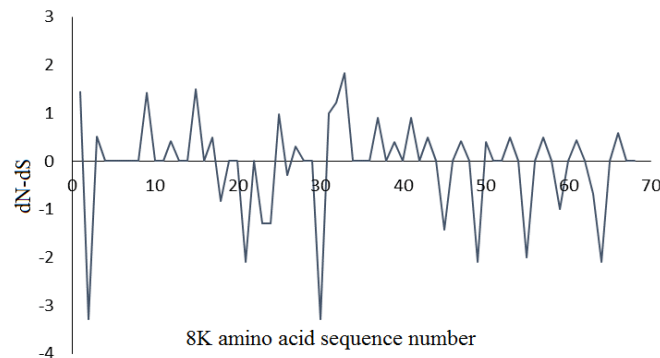
Figure 2: The dN-dS along 72 samples of 8K amino acid sequences. This ratio is achieved by differences between nonsynonymous and synonymous positions (dN-dS). The dN-dS index below zero is consistent with negative selection against protein change. In contrast, a dN-dS index above zero may be an indication that adaptive or positive selection is driving gene divergence. dS and dN are the numbers of synonymous and nonsynonymous substitutions per site, respectively.
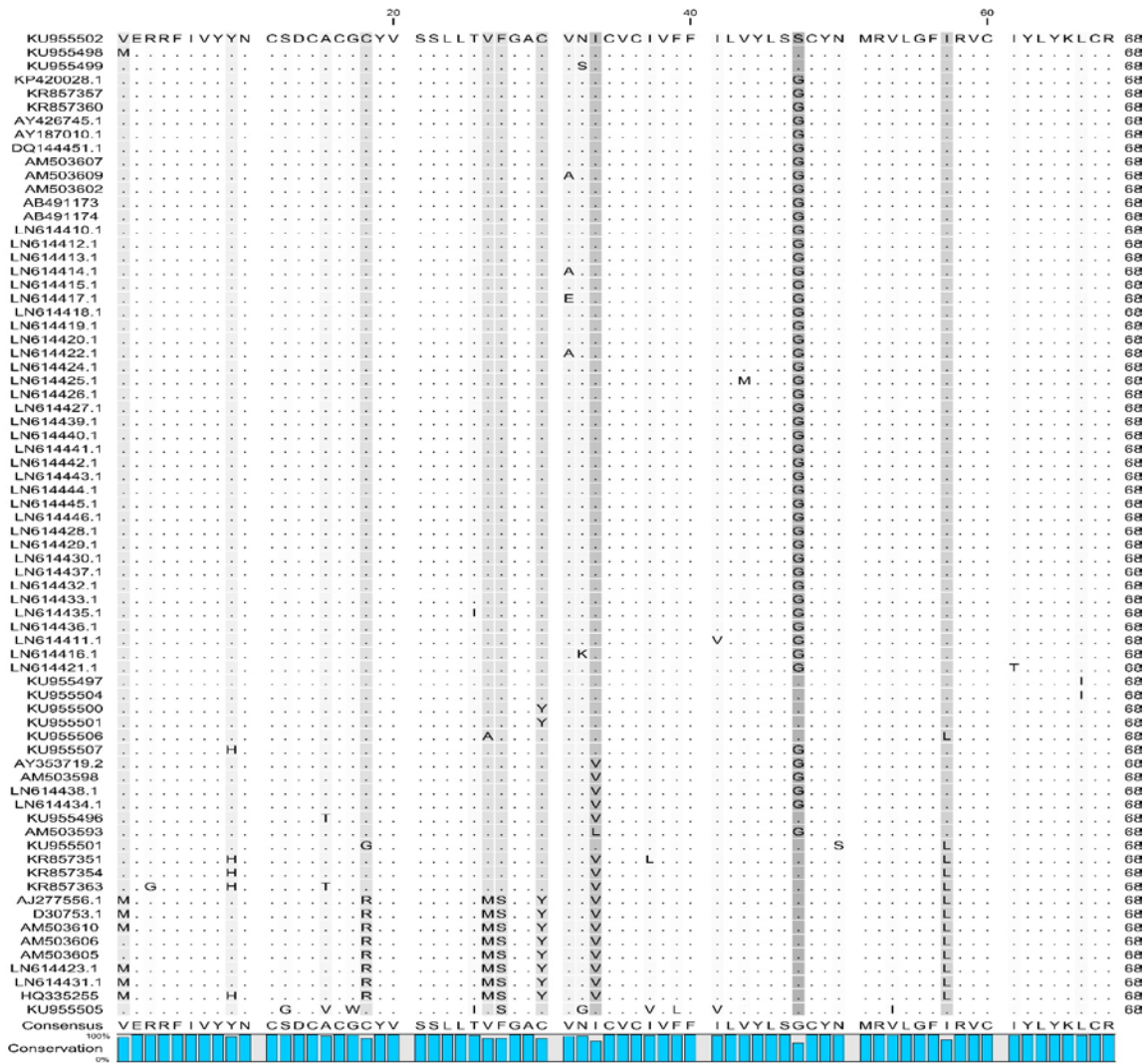
Figure 3: Multiple alignment of the amino-acid sequences of PMTV 8K suppressors from the worldwide isolates.

In this study, we used Shannon entropy-based approach to evaluate the variability of each site of nucleotide and corresponding amino acid. Based on Shannon entropy method, 139 and 97 nucleotide sites had the highest entropy value, while 47 and 33 amino acid sites showed the most diversity along 8K sequences (Figs. 4, 5).
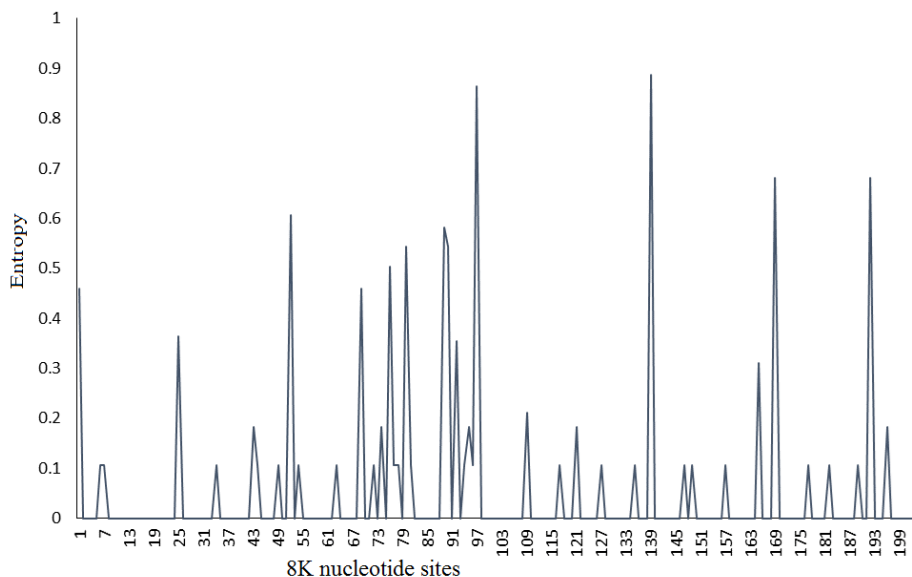

Figure 4: Shannon entropy for 8K nucleotide sites from 72 worldwide sequences. Entropy was calculated for each nucleotide position along 8K.
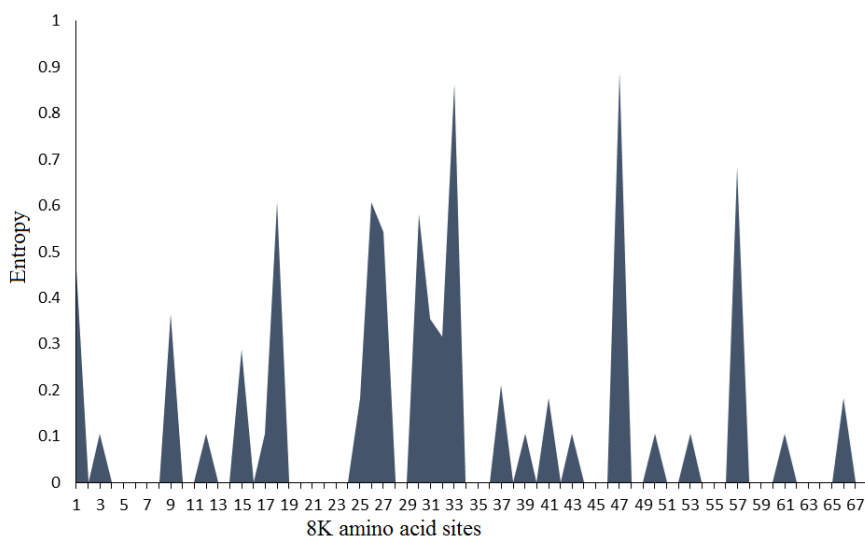

Figure 5: Histograms of amino acid positions occurring at different entropy values in 8K genetic diversity study.

DISCUSSION

RNA viruses are faced with different host antiviral responses including RNA silencing that performs sequence-specific inhibition of viral mRNAs expression. To counteract the host RNA silencing, viruses develop VSRs such as the 8K protein of PMTV. In this study, the complexity of 8K nucleotide and amino acid sites was characterized. Position 33 indicated the highest

evolutionary rate based on the mean evolutionary rate for each site, which was further confirmed by Shannon entropy approach (with a high-value complexity) and strongly positive selection. The evolutionary rate results were confirmed by the *dN-dS* index, where a strongly positive selection was observed on position 33. Therefore, this amino acid evolves more than those in the other positions. Furthermore, the most positive pressure on position 33 shows its importance for virus survival and flexibility in response to evolutionary forces. While, the lowest evolutionary rates were related to the amino acid positions 21, 22 and 46. We assume that these amino acid sites can play an important role in 8K protein function. The results derived here suggested that Shannon's information entropy can be used as a measure to evaluate the genetic diversity of viruses and their evolution [22, 23]. However, there is increasing tendency to incorporate entropy methods into population genetic analysis that it might be extended to virus diversity and evolution. Twenty-one positively selected sites were observed during the 8K sequence, which represents that the *8K* gene is under strong positive selection suggesting the acceleration in the divergence of the *8K* gene. Recently, positive selection on another RNA silencing suppressor, P1 of Rice yellow mottle virus has been reported [27]. Thus, the increased rate of VSRs evolution might be due to reduced constraint than other viral genes with highly conserved patterns [20]. Moreover, our results showed that the AUG start codon can be replaced by GUG in *8K* of PMTV isolates. The utilization of initiating non-AUG codon for maintenance of proper ratios of proteins has been reported for both cellular and viral mRNAs [28], including plant viruses [29]. The increase of GUG start codon in PMTV isolates can lead to change in expression level of 8K protein. Further studies would be concentrated on assessing the impact of AUG and GUG start codons on the virulence of the virus.

Point mutations on VSRs, including the helper component proteinase (HC-Pro), affect their suppression function [30-32]. Also, mutations on sites under positive selection can enhance the strength of 8K suppressor. Another important mechanism of the 8K evolution and virus genetic variation is recombination, which can also play a key role in maintaining the effectiveness of purifying selection by preventing the accumulation of deleterious mutations [33]. The high number of sites positively selected and recombination events suggest that 8K suppressor evolves rapidly. Also, negative Tajima's D showed demographic forces and expansion on PMTV population. Thus, the variability and acceleration in the divergence of the *8K* gene might be a virus strategy to constitute an adaptation to the different plant hosts in various agro-systems during PMTV evolution. The rapid evolution of the PMTV 8K could be due to the evolutionary conflict between the RNA silencing pathway and its suppression to increase virus replication and spread [34].

In this study, we determined the high evolutionary rate of 8K sequences using bioinformatics tools. Considering the role of 8K in enhancing the virulence of PMTV and also as an RNA silencing suppressor of plants, its variability might be possible to play a big threat to infect other plants or overcome resistant cultivars.

Conflict of Interest: The authors declare that they have no conflict of interest.

REFERENCES

1. Adams MJ, Antoniw JF, Kreuze J. Virgaviridae: a new family of rod-shaped plant viruses. Arch Virol 2009;154:1967-1972.

2. Hu X, Dickison V, Lei Y, He C, Singh M, Yang Y, Xiong X, Nie X. Molecular characterization of Potato mop-top virus isolates from China and Canada and development of RT-PCR differentiation of two sequence variant groups. Can J Plant Pathol 2016;38:231-242.

3. Ramesh SV, Raikhy G, Brown CR, Whitworth JL, Pappu HR. Complete genomic characterization of a Potato mop-top virus isolate from the United States. Arch Virol 2014; 159:3427-3433.

4. Santala J, Samuilova O, Hannukkala A, Latvala S, Kortemaa H, Beuch U, Kvarnheden A, Persson P, Topp K, Ørstad K, Spetz C. Detection, distribution and control of potato mop-top virus, a soil-borne virus, in northern Europe. Ann Appl Biol 2010;157:163-178.

5. Scott KP, Kashiwazaki S, Reavy B, Harrison BD. The nucleotide sequence of potato mop-top virus RNA 2: a novel type of genome organization for a furovirus. J Gen Virol 1994;75: 3561- 3568.

6. Savenkov EI, Sandgren M, Valkonen JP. Complete sequence of RNA 1 and the presence of tRNA-like structures in all RNAs of *Potato mop-top virus*, genus *Pomovirus*. J Gen Virol 1999;80:2779-2784.

7. Sandgren M, Savenkov EI, Valkonen JP. The readthrough region of *Potato mop-top virus* (PMTV) coat protein encoding RNA, the second largest RNA of PMTV genome, undergoes structural changes in naturally infected and experimentally inoculated plants. Arch Virol 2001;146:467-477.

8. Morozov SY, Solovyev AG. Triple gene block: modular design of a multifunctional machine for plant virus movement. J Gen Virol 2003;84:1351-1366.

9. Zamyatnin Jr AA, Solovyev AG, Savenkov EI, Germundsson A, Sandgren M, Valkonen JP, Morozov SY. Transient coexpression of individual genes encoded by the triple gene block of Potato mop-top virus reveals requirements for TGBp1 trafficking. Mol Plant Microbe Interact 2004;17:921-930.

10. Lukhovitskaya NI, Thaduri S, Garushyants SK, Torrance L, Savenkov EI. Deciphering the mechanism of defective interfering RNA (DI RNA) biogenesis reveals that a viral protein and the DI RNA act antagonistically in virus infection. J Gen Virol 2013;87:6091-6103.

11. Nielsen SL, Nicolaisen M. Identification of two nucleotide sequence sub-groups within *Potato mop-top virus*. Arch Virol 2003;148:381-388.

12. Xu H, DeHaan TL, De Boer SH. Detection and confirmation of *Potato mop-top virus* in potatoes produced in the United States and Canada. Plant Dis 2004;88:363-367.

13. Cerovska N, Pecenkova T, Filigarova M, Dedic P. Sequence analysis of the Czech potato mop-top virus (PMTV) isolate Korneta-Nemilkov. Folia Microbiol 2007;52:61-64.

14. Latvala-Kilby S, Aura JM, Pupola N, Hannukkala A, Valkonen JP. Detection of Potato mop-top virus in potato tubers and sprouts: combinations of RNA2 and RNA3 variants and incidence of symptomless infections. Phytopathology 2009;99:519-531.

15. David N, Mallik I, Crosslin JM, Gudmestad NC. First report of *Potato mop-top virus* in North Dakota. Plant Dis 2010;94:1506.

16. Whitworth JL, Crosslin JM. Detection of *Potato mop top virus* (*Furovirus*) on potato in southeast Idaho. Plant Dis 2013;97:149.

17. Brigneti G, Voinnet O, Li WX, Ji LH, Ding SW, Baulcombe DC. Viral pathogenicity determinants are suppressors of transgene silencing in *Nicotiana benthamiana*. EMBO J 1998;17:6739-6746.

18. Voinnet O, Pinto VM, Baulcombe DC. Suppression of gene silencing: a general strategy used by diverse DNA and RNA viruses of plants. Proc Natl Acad Sci USA 1999;96:14147-14152.

19. Pumplin N, Voinnet O. RNA silencing suppression by plant pathogens: defence, counter-defence and counter-counter-defence. Nat Rev Microbiol 2013;11:745.

20. Murray GGR, Kosakovsky Pond SL, Obbard DJ. Suppressors of RNAi from plant viruses are subject to episodic positive selection. Proc Royal Soc 2013;280:20130965.

21. Kalyandurg P, Gil JF, Lukhovitskaya NI, Flores B, Müller G, Chuquillanqui C, Palomino L, Monjane A, Barker I, Kreuze J, Savenkov EI. Molecular and pathobiological characterrization of 61 Potato mop-top virus full-length cDNAs reveals great variability of the virus in the center of potato domestication, novel genotypes and evidence for recombination. Mol Plant Pathol 2017;18:864-877.

22. Kosman E, Leonard KJ. Conceptual analysis of methods applied to assessment of diversity within and distance between populations with asexual or mixed mode of reproduction. New Phytol 2007;174:683-696.

23. Sherwin WB. Entropy and information approaches to genetic diversity and its expression: genomic geography. Entropy 2010;12:1765-1798.

24. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol 2016;33:1870-1874.

25. Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 2003;19:2496-2497.

26. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci: CABIOS 1992;8:275-282.

27. Sereme D, Lacombe S, Konate M, Bangratz M, Pinel-Galzi A, Fargette D, Traoré AS, Konaté G, Brugidou C. Sites under positive selection modulate the RNA silencing suppressor activity of rice yellow mottle virus movement protein P1. J Gen Virol 2014;95: 213-218.

28. Touriol C, Bornes S, Bonnal S, Audigier S, Prats H, Prats AC, Vagner S. Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons. Biol Cell 2003;95:169-178.

29. Shirako Y. Non-AUG translation initiation in a plant RNA virus: a forty-amino-acid extension is added to the N terminus of the soil-borne wheat mosaic virus capsid protein. J Virol 1998;72:1677-1682.

30. González-Jara P, Atencio FA, Martínez-García B, Barajas D, Tenllado F, Díaz-Ruíz JR. A single amino acid mutation in the Plum pox virus helper component-proteinase gene abolishes both synergistic and RNA silencing suppression activities. Phytopathology 2005; 95:894-901.

31. Yambao MLM, Yagihashi H, Sekiguchi H, Sekiguchi T, Sasaki T, Sato M, Atsumi G, Tacahashi Y, Nakahara KS, Uyeda I. Point mutations in helper component protease of clover yellow vein virus are associated with the attenuation of RNA-silencing suppression activity and symptom expression in broad bean. Arch Virol 2008;153:105-115.

32. Torres-Barceló C, Martín S, Daròs JA, Elena SF. From hypo-to hypersuppression: effect of amino acid substitutions on the RNA-silencing suppressor activity of the Tobacco etch potyvirus HC-Pro. Genetics 2008;180:1039-1049.

33. Fraile A, Alonso-Prados JL, Aranda MA, Bernal JJ, Malpica JM, Garcia-Arenal F. Genetic exchange by recombination or reassortment is infrequent in natural populations of a tripartite RNA plant virus. J Virol 1997;71:934-940.

34. Obbard DJ, Jiggins FM, Halligan DL, Little TJ. Natural selection drives extremely rapid evolution in antiviral RNAi genes. Curr Biol 2006;16:580-585.