Plant Pathol. J. 32(1) : 33-46 (2016)
http://dx.doi.org/10.5423/PPJ.OA.07.2015.0144
pISSN 1598-2254 eISSN 2093-9280

The Plant Pathology Journal
©The Korean Society of Plant Pathology

Research Article Open Access

Population Analysis of Iranian *Potato virus Y* Isolates Using Complete Genome Sequence

Reza Pourrahim* and Shirin Farzadfar
Plant Virus Research Department, Iranian Research Institute of Plant Protection (IRIPP), P.O. Box 19395-1454. Tehran, Iran

(Received on July 25, 2015; Revised on September 21, 2015; Accepted on October 12, 2015)

In this study, the full-length nucleotide sequences of four Iranian PVY isolates belonging to PVY^N strain were determined. The genome of Iranian PVY isolates were 9,703–9,707 nucleotides long encoding all potyviral cistrons including P1, HC-Pro, P3, 6K1, CI, 6K2, VPg, NIa-Pro, NIb and CP with coding regions of 825, 1,395, 1,095, 156, 1,902, 156, 564, 732, 1,557 and 801 nucleotides in length, respectively. The length of pipo, embedded in the P3 cistron, was 231 nucleotides. Phylogenetic analysis showed that the Iranian isolates clustered with European recombinant NTN isolates in the N lineage. Recombination analysis demonstrated that Iranian PVY^N isolates had a typical European PVY^{NTN} genome having three recombinant junctions while PVY^N and PVY^O were identified as the parents. We used dN/dS methods to detect candidate amino acid positions for positive selection in viral proteins. The mean ω ratio differed among different genes. Using model M0, ω values were 0.267 (P1), 0.085 (HC-Pro), 0.153 (P3), 0.050 (CI), 0.078 (VPg), 0.087 (NIa-pro), 0.079 (NIb) and 0.165 (CP). The analysis showed different sites within P1, P3 and CP were under positive selection pressure, however, the sites varied among PVY populations. To the best of our knowledge, our analysis provides the first demonstration of population structure of PVY^N strain in mid-Eurasia Iran using complete genome sequences and highlights the importance of recombination and selection pressure in the evolution of PVY.

Keywords **:** phylogeny, PVY^N, recombination, variation

*Corresponding author.
Phone) +98-21-22403012, FAX) +98-21-22402570
E-mail) pourrahim@yahoo.com
This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Potato virus Y (PVY; genus *Potyvirus*, family *Potyviridae*) is one of the most widespread and important viruses infecting agriculturally important plants belonging to the family *Solanaceae* including potato, tobacco, and pepper (De Bokx and Huttinga, 1981; Shukla et al., 1994). PVY is transmitted by at least 40 aphid species (family *Aphididae*) in a nonpersistent manner (Edwardson and Christie, 1997). Vertical transmission is also a serious problem for commercial tuber seed production in potato (Blanc et al., 1997). The monopartite genome is composed of a single-stranded, positive-sense RNA molecule of about 10,000 nucleotides long containing a single open reading frame (ORF) (Lorenzen et al., 2006). The encoded large polyprotein is processed by three virus-encoded proteases to produce at least 10 mature functional proteins (Urcuqui-Inchima et al., 2001).

Several distinct strains of PVY including tobacco veinal necrosis PVY^N (N strain) and tuber necrosis strain PVY^{NTN}, which induces potato tuber necrotic ringspot disease (PTNRD), have been recognized (Tian et al., 2011). Isolates of the N strain are not harmful only to tobacco but are also problematic in potato, where they overcome *Ny* and *Nc* genes conferring hypersensitive resistance to PVY isolates belonging to the strain groups PVY^O and PVY^C, respectively (Singh et al., 2008). Despite attempts to limit their spread (Singh, 1992), PVY^N isolates are now distributed all over the world and are of a matter of great concern to growers because of frequent emergence of new variants via mutation and/or recombination with other PVY isolates (Singh et al., 2008).

Studies of the molecular evolutionary history of viruses are very complex as they involve understanding variation caused by mutation, recombination, selection pressure and adaptation (Moury et al., 2004; Moury et al., 2002). These studies help to understand important aspects of viral biology such as changes in virulence and geographical ranges that leads to 'emergence' of new epidemics. Hence, information on this aspect of viruses is essential to design

management and control strategies in order to hinder their spread.

Potato (*Solanum tubersum* L.) is an agriculturally and economically important crop in Iran that is grown over 190,000 hectares with an estimated production of 5.5 Million tons (FAO, 2013). PVY is one of the most damaging potato viruses in Iran and is responsible for the main yield losses (Pourrahim et al., 2007). Numerous studies have focused on etiology, pathogenesis, ecology, molecular biology and control of the virus (Hosseini et al., 2011; Pourrahim et al., 2007); however, there are still gaps in our knowledge in terms of genetic variability and population structure especially at full-length genome level. In the current study, we used different experimentally evolutionary approaches in order to understand the factors that influence PVY variability. By measuring the selective pressure exerted on the proteins encoded by the PVY genome, we found that mutations resulting in amino acid substitutions also contribute to PVY evolution. This analysis rarely done on plant viruses can also help protein function and structure studies as well as virus epidemiology.

Materials and Methods

Virus isolates. Forty five Iranian PVY isolates were selected from our previous surveys conducted in Iran (Pourrahim et al., 2007) including 14 new isolates that were collected during 2009 to 2013 (unpublished data). The new isolates were tested for the presence of PVY by TAS-ELISA using PVY^N, $PVY^{O/C}$ and PVY^O-specific monoclonal antibodies (Neogen, UK). Isolates were assigned a subscript of O, N, C or NTN depending on the serological data. All samples were homogenized in 0.1 M K-phosphate buffer (pH 7.4) and mechanically inoculated onto tobacco (*Nicotiana tabacum* cvs. Samsun or Xanthi).

RT-PCR, cloning and sequencing. Details of the isolates including original host plants and the year of isolation are shown in Supplementary Table S1. The full-length genomic sequences of four PVY^N strain (IRNH1, IRNH172, IRND41, IRNZ33) were amplified by RT-PCR from the total RNA isolated from systemically infected *N. tabacum* cv. Xanthi leaves using an RNeasy Plant Mini Kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions. The RNAs were reverse transcribed and PCR amplified using high-fidelity PlatinumTM *Pfu* DNA polymerase (Invitrogen, Carlsbad, CA, USA). The 5' terminus of the PVY genome was obtained using a 5' RACE kit (Roche Diagnostics, Mannheim, Germany) according to the manufacturer's instructions. The resultant amplicons were purified using the QIAquick Gel Extraction Kit (QIAGEN, Hilden, Germany) and subjected to direct sequencing or cloned into *Eco*RV site of plasmid pZErO-2 (Invitrogen, Carlsbad, CA, USA), transferred into *E. coli* strain DH5α and plated on media containing 25 μg ml^{-1} kanamycin. Plasmids were purified using QIAprep Spin Miniprep Kit (QIAGEN, Hilden, Germany). PCR products or cloned fragments were sequenced by a primer walking approach in both directions using the BigDye Terminator version 3.1 Cycle Sequencing Ready Reaction Kit (Applied Biosystems) and an Applied Biosystems Genetic Analyser DNA Model 310. For sequencing of each PCR fragment different forward and reverse primers with 400 to 450 nt lengths apart were designed (data not shown) and sequences overlapping by at least 100 bp were obtained. Sequence data were edited and assembled using BIOEDIT version 5.0.9 (Hall, 1999).

Detection of recombination. The genomic sequences of PVY^N isolates that were obtained in this study and those of 54 full-length sequences obtained from GenBank were used for recombination analyses (Table S1). Relationships of aligned polyprotein sequences were calculated separately using Neighbor-Joining (NJ) and Maximum Likelihood (ML) methods. Amino acid sequences were aligned using CLUSTAL X2 with TRANSALIGN for optimal alignment (Weiller, 1998). Recombination events, putative parental isolates of recombinants, and recombination break points were analyzed using several methods implemented in the RDP3 version 3.44b (Martin et al., 2010) with default configuration and Bonferroni corrected *P*-value cut-off of 0.05 and 0.01. Only those potential recombination events detected by at least three different methods with an associated P value of 1.0×10^{-6} were considered, and involving fragments sharing $\geq 97\%$ sequence identity with their parental sequences were considered. All putative recombinants identified by the RDP3 program were rechecked using SISCAN version 2 (Gibbs et al., 2000).

Phylogenetic relationship and genetic variability. Phylogenetic relationship was inferred using the NJ, and ML methods implemented in the MEGA 5 (Tamura et al., 2011) and ML tree is shown (Fig. 3). We used *Pepper mottle virus* (PeMoV) (GenBank accession no. NC_001517) as an outgrop. Pairwise genetic distances were analyzed by the Kimura's two-parameter method using MEGA v. 4.1 program (Tamura et al., 2007). DNASP version 4.10 (Rozas et al., 2003) was used to estimate haplotype diversity which was calculated based on the frequency and number of haplotypes in the population.

Analyses of selection pressures. The direction and strength of selection is measured by ω, the nonsynonymous to synonymous substitution rate ratio (dn/ds = ω), with $\omega < 1$, = 1, and > 1 indicating purifying selection, neutral evolution, and positive selection, respectively. The non-synonymous to synonymous nucleotide substitution rate ratio (ω) was assessed using an ML codon-substitution model implemented in the CODEML program of the PAML4 package (Yang, 2007). Models M0, M1 and M7 do not allow for the existence of positively selected sites. M0 calculates a single ω ratio (between 0 and 1) averaged over all sites, M1a (nearly neutral) account for neutral evolution by estimating the proportion of conserved ($\omega = 0$) and neutral ($\omega = 1$) sites, and M7 (β) uses a discrete beta distribution between 0 and 1 to model different ω ratios between sites. Three models M2a (positive selection), M3 (discrete), and M8 (β plus ω) account for positive selection by using parameters that can estimate $\omega > 1$ (Wong et al., 2004; Yang et al., 2000; Yang et al., 2005). The first step in identifying amino acid sites under positive selection is to test whether sites exist with $\omega > 1$ by application of likelihood ratio test (LRTs) to compare nested models, therefore, three LRTs (M3 vs M0, M2a vs M1a and M8 vs M7) were used to assess the models' fit to the data, as described by Wong et al. (2004). Where the LRTs suggested positive selection, the Bayesian methods (Bayes Empirical Bayes – BEB) approach (Yang et al., 2000) was used to identify amino acids subjected to positive selection. Sites having high posterior probabilities ($P > 90\%$) of belonging to the site class with $\omega > 1$ are good candidates for positively selected sites.

Results

Biological and molecular characterizations of the Iranian PVY^N isolates. Four isolates out of fifteen positive samples identified by monoclonal antibody against PVY^N

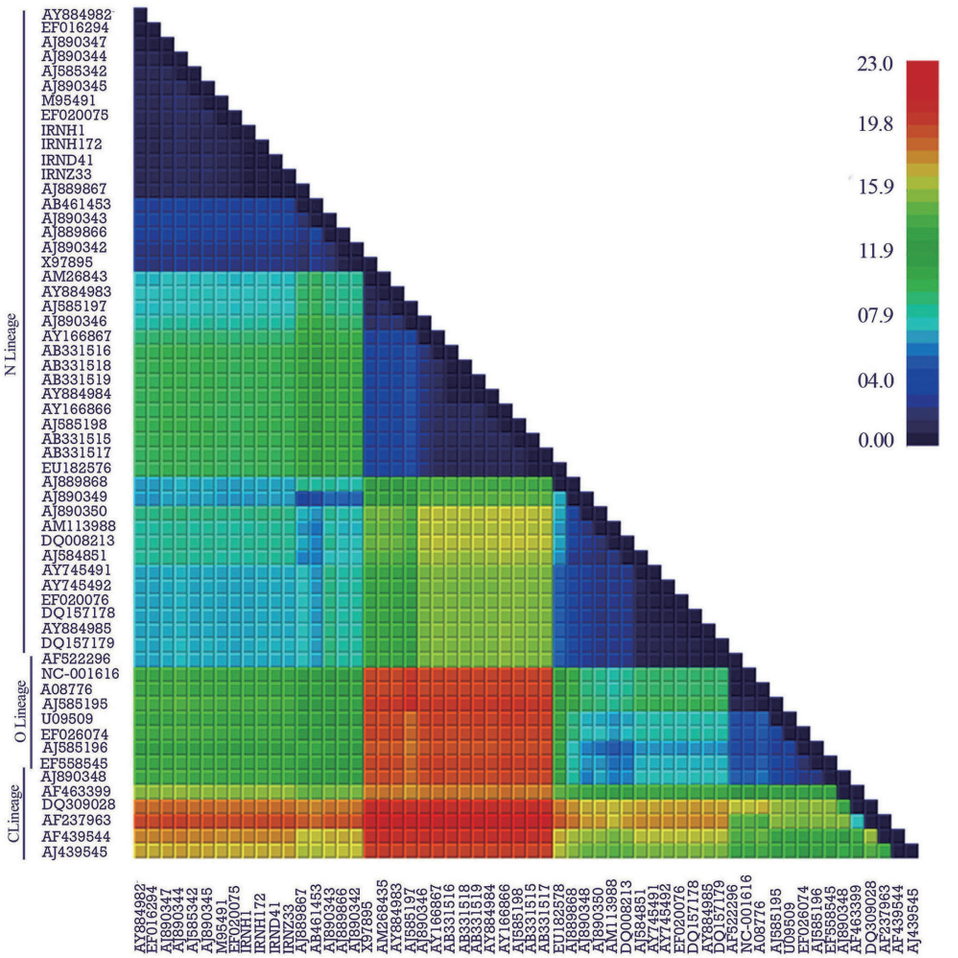

Fig. 1. Two dimensional nucleotide diversity plot constructed using complete genome sequences of four PVY^N Iranian isolates (IRNH1, IRNH172, IRND41 and IRNZ33) with 54 others one obtained from GenBank.

were selected. All of the samples induced vein necrosis in *N. tabacum* cv. Xanthi. The genomes of selected Iranian isolates were 9703–9707 nucleotides long. For recombination analysis, the terminal primer sequences including 5'UTR (189 nucleotides) and 3'UTR (328 to 332 nucleotides) were omitted. All of the motifs reported for different potyvirus genes and encoded proteins were found. The regions encoding the P1 (190–1014 nucleotides), helper-component proteinase (HC-Pro) (1015–2409 nucleotides), P3 (2410–3504 nucleotides), 6 kDa 1 protein (6K1) (3505–3660 nucleotides), cylindrical inclusion protein (CI) (3661–5562 nucleotides), 6 kDa 2 protein (6K2) (5563–5718 nucleotides), genome-linked viral protein (VPg) (5719–6282 nucleotides), nuclear inclusion a-proteinase (NIa-Pro) (6283–7014 nucleotides), nuclear inclusion b (NIb) (7015–8571 nucleotides) and coat protein (CP) (8572–9376 nucleotides) were 825, 1395, 1095, 156, 1902, 156, 564, 732, 1557 and 801 nucleotides long, respectively. Furthermore, the region from 2919 to 3149 (231 nucleotides) was determined for *pipo* ORF. Genome-wide pairwise nucleotide diversity analysis of the complete genome revealed three regions representing: (0.0–11.9), (11.9–15.9) and (15.9–23.0) nucleotide diversity for N, O and C lineages, respectively. The lowest nucleotide diversity (0.0–4.0) was found for N-Europe followed by 4.0–7.9 for W (PVY^{N-Wi}) and 7.9–11.9 for N-Europe & North America populations from N lineage. The highest nucleotide diversities were found for O and C lineages (Fig. 1).

Recombination analysis. Different methods were used for recombination breakpoint prediction and provided evidence for clear recombination (Fig. 2, Table 1). Remarkably, following the adopted criteria (detectable by at least three different methods and recombined fragments with ≥ 97% nucleotide identity with parental sequences), recombination events observed across all four Iranian viral genomes analyzed. Recombination analysis demonstrated that Iranian PVY^{N} had a typical European PVY^{NTN} genome having

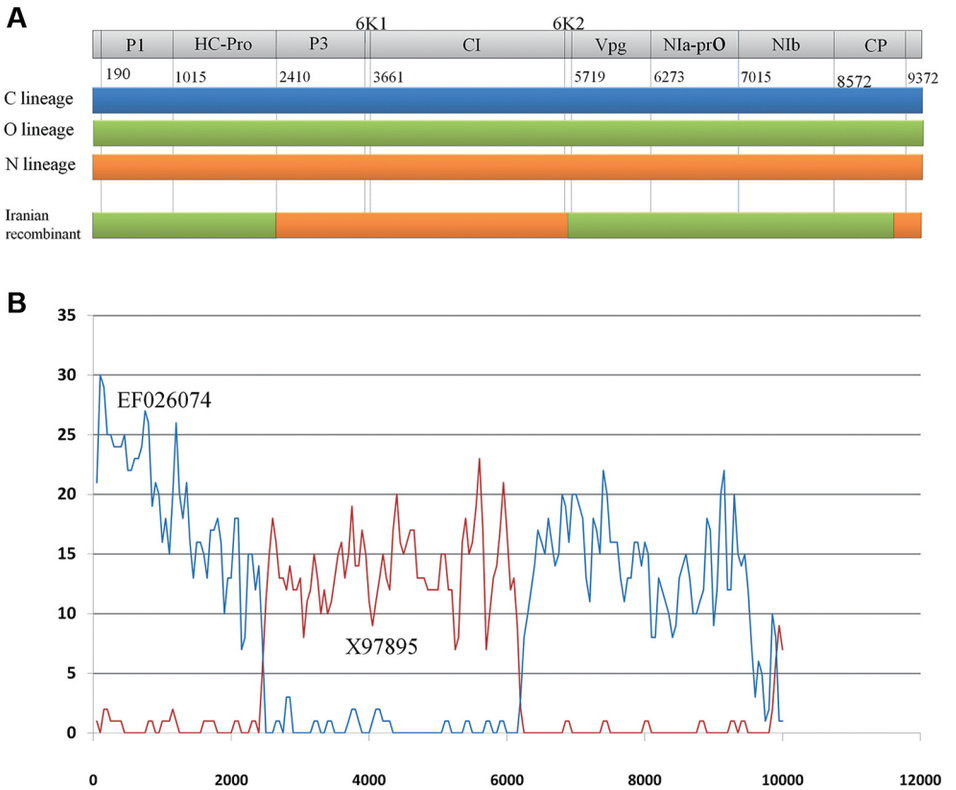

Fig. 2. Recombination pattern of *Potato virus Y* genome. (A) Color figure of each lineage and recombination pattern of Iranian isolates. (B) Graphs showing SISCAN Version 2 analysis of the full-length sequence of Iranian isolate (IRNH172) with that of EF026074 (Blue line) as well as X97895 (Red line). X and Y axes represent *Z*-value and genome nucleotides, respectively. The sequences of EF026074 and X97895 represent the likely parental sequences of IRNH172. Iranian recombinant isolate (IRNH172) was more closely related to EF026074 between approximately 190 to 2,244 nt, and 8,977 to 9,703 nt; and more closely related to X97895 between 5,581 to 8,977 nt (i.e. *Z*-value > 3.0).

Table 1. Recombination crossover sites in *Potato virus Y* isolates detected by recombination detecting programs

Recombinant	Event	Parental isolates[a]		Breakpoints[c]		Methods[d]	SISCAN (*Z*-value)[e]
		Major parent (% identity)[b]	Minor parent (% identity)	Begin	End		
Iranian isolates	1	EF026074 (O) (92.0)	X97895 (N) (99.0)	P1 / 190	HC-Pro / 2244	RGBMCS	12.0
	2	EF026074 (O) (92.0)	X97895 (N) (99.0)	6K2 / 5581	CP / 8977	RGBMCS	10.0
	3	EF026074 (O) (92.0)	X97895 (N) (99.0)	CP / 8977	CP / 9703	RGBMCS	8.0

[a]Parental isolate means the most likely isolate among analyzed isolates.
[b]Nucleotide sequence identity in particular genome region of each lineage is indicated in parentheses.
[c]Numbers indicate recombination sites.
[d]R, G, B, M, C, and S indicate detection by RDP, GENCONV, BOOTSCAN, MAXCHI, CHIMAERA and SISCAN methods, respectively, with the presented highest *P*-value being that determined by the method indicated in bold type.
[e]The *Z*-value was calculated by original SISCAN v.2 program.

three recombinant junctions, while PVY^N and PVY^O were identified as the parents. The recombination breakpoints were detected between P1 and HC-Pro (at nucleotide positions 190 and 2244), 6K2 and CP (at nucleotide positions 5581 and 8977), and in the CP (nucleotide positions 8977 and 9703) (Table 1). Every event identified by the RDP3 program was confirmed by SISCAN V2 with parental-like Oz (a Canadian isolate with accession no. EF026074) and a Switzerland isolate CH-605(NR) (accession no. X97895) from O and N lineages, respectively (Fig. 2A and B).

Phylogenetic relationships. Pairwise comparisons using CLUSTAL X2 were performed and after being degaped, there were a total of 9165 positions in the final dataset. Phylogenetic trees showed that PVY isolates fell into three main lineages where four Iranian PVY isolates (IRNH1, IRNH172, IRND41, IRNZ33) were clustered with European NTN strain (Fig. 3). The pairwise genetic diversity within population were 0.074 ± 0.002, 0.035 ± 0.001 and 0.120 ± 0.003 for N, O and C lineages, respectively (Table 2). The highest pairwise diversity for C lineage may indicate the isolates in this lineage were more genetically diverse than N and O lineages. In addition, the between-population diversities (Table 3) were identified that were different to the within population diversity, indicating that there was genetic differentiation between these populations. The highest nucleotide diversities 0.229 ± 0.004 were determined between C and N-Europe populations (Table 3).

Selection analyses. The mean ω ratio differed greatly among different cistrons. Using model M0, ω values were 0.267 (P1), 0.085 (HC-Pro), 0.153 (P3), 0.050 (CI), 0.078 (VPg), 0.087 (NIa-pro), 0.079 (NIb) and 0.165 (CP) (Table 4). Model M0, used to assess selection pressures [maximum likelihood (ML) framework of codon substitution], and yielded the value ≥ 0.267 over all codon sites in the 10 cistrons indicating purifying selection (Table 4). For P1 protein, heterogeneity of selection pressure, tested by using M3 vs M0 (LRT), revealed that M3 fitted the data significantly better than M0. Model M3 suggested that a large set of sites (56.2%) were evolving under strong purifying selection (ω = 0.062), fewer sites (31.6%) under weak purifying selection (ω = 0.441) and only 1.2% of sites under positive selection (ω = 1.131). Comparison of M8 with M7 showed that 32 sites (14.3%) were under positive selection (ω = 1.047) (Table 4). By using the BEB inference (Yang et al., 2005) 22 sites of P1 protein were identified under positive selection. Additionally the Naive Empirical Bayes (NEB) analysis inference (Yang et al., 2000) under M8 provided 10 more sites for P1 protein under positive selection (Table 4). For HC-Pro, M3 suggested that a large set of sites (69.5%) were evolving under strong purifying selection pressure (ω = 0.010), 28.6% were under weak purifying selection pressure (ω = 0.237) and only 1.7% of sites were under positive selection pressure (ω = 0.829). Using the NEB inference five sites of HC-Pro identified under positive selection pressure with models M3 and M8 but the LRT statistics for these sites were not significant. For P3, 53.1% of sites were evolving under strong negative selection pressure (ω = 0.006), 36.5% of sites under purifying selection pressure and 1.03% were under positive selection pressure. Using M3, M2a and M8, BEB predicted the 35 amino acids residue under positive selection (Table 4). For CI and VPg proteins the majority of sites were found under strong negative selection pressure with ω = 0.014 and ω = 0.020, respectively. However NEB and BEB analysis identified some sites under positive selection pressure with M3 and M8, respectively (Table 4). Although a few sites were found under positive selection pressure for NIa and NIb proteins using M8; however, the LRT statistics for these

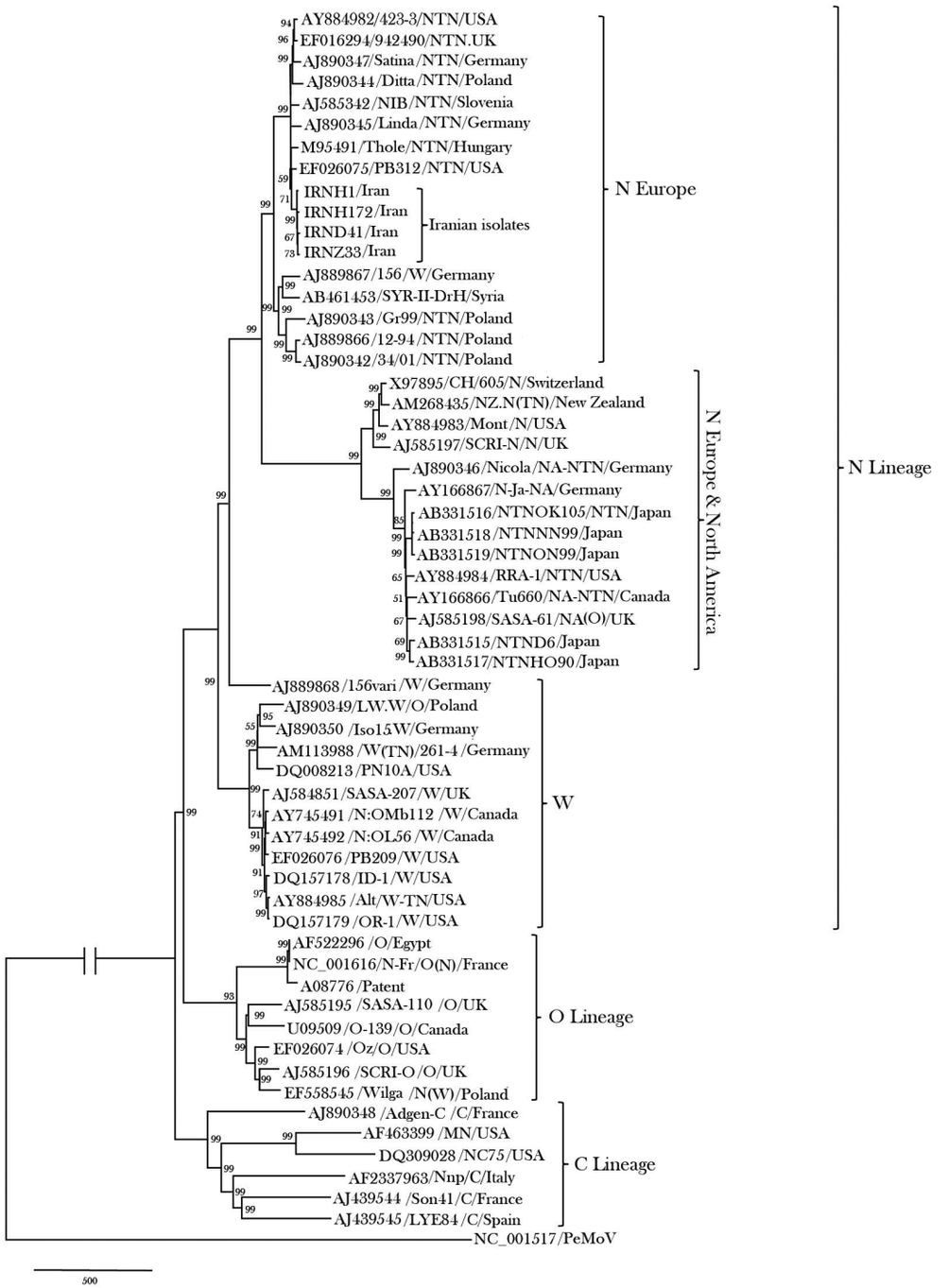

Fig. 3. A maximum likelihood (ML) phylogenetic tree showing the relationship among *Potato virus Y* isolates. The tree was constructed from the full-length sequences of 58 PVY isolates, including recombinant isolates. Numbers at each node indicate the percentage of supporting puzzling steps (or bootstrap samples) in ML method. The name of each isolate, its country of origin, strain relationship, and GenBank accession number are listed. We used *Pepper mottle virus* (PeMoV) (accession no. NC_001517) as an outgrop.

proteins were not significant.

Comparison of M8 with M7 for CP showed that M8 fitted the data better than M7 and 16 sites were under positive selection pressure. Seven sites (Table 4) identified by using the BEB inference with probabilities of 95 to 99% and high ω ratio (ω = 1.369); however, for nine sites the LRT statis-

Table 2. Haplotype and nucleotide diversity of each *Potato virus Y* population

Lineage		Haplotype diversity	Nucleotide diversity
Sub-populations	N	0.998 (0.000)	0.074 (0.002)
	N-Europe	1.000 (0.003)	0.015 (0.001)
	N-Europe & North America	0.989 (0.000)	0.022 (0.001)
	W (PYV^{N-Wi})	1.000 (0.002)	0.016 (0.001)
O		0.964 (0.005)	0.035 (0.001)
C		1.000 (0.003)	0.120 (0.003)

Nucleotide diversity was estimated by the average pairwise difference among sequence in a sample, based on all sites (Tamura et al., 2007). Numbers in parentheses indicate standard deviations.

Table 3. Mean nucleotide diversity between *Potato virus Y* populations

Populations	C	O	N-Europe	N-Europe & North America	PYV^{N-Wi}
C	–	0.138 (0.003)	0.229 (0.004)	0.111 (0.002)	0.159 (0.003)
O	–	–	0.128 (0.003)	0.187 (0.004)	0.075 (0.002)
N-Europe	–	–	–	0.089 (0.002)	0.072 (0.003)
N-Europe & North America	–	–	–	–	0.151 (0.004)
W (PYV^{N-Wi})	–	–	–	–	–

Nucleotide diversity was estimated by the average pairwise difference among sequence in a sample, based on all sites (Tamura et al., 2007). Numbers in parentheses indicate standard deviations.

tics were not significant (Table 4). No positive sites were found for 6K1 and 6K2 proteins under different models. Comparison of P3 amino acid sites between PVY lineages showed that most sites under positive selection belonged to

Table 4. The dn/ds (ω) values, log-likelihood (ln*L*) values, likelihood ratio test (LRT) statistics and positively selected amino acid sites undergoing different models of codon substitution were used to investigate selection pressures on ten proteins encoded by the *Potato virus Y* genome analysed in this study

Protein	Models[1] / Site models	Parameter estimates	ω ratio	ln*L*	LRT[2]	Positively selected (amino acids) sites[3]
P1	M0	ω=0.267	0.267	-6441.013		None
	M3	$p0$=0.562, $p1$=0.316, $p2$=0.120, $\omega0$=0.062, $\omega1$=0.441, $\omega2$=1.131	0.311	-6314.060	P < 0.01 (0.0)	26 L, 34 S, 36 T, 52 R, 55 E, 56 F, 67 C*, 78 A, 116 L, 134 A, 137 Y, 138 H, 140 P, 184 H, 185 R*, 187 V, 203 R, 206 K, 209 V, 210 V, 211 R*, 212 L, 213 Q, 214 H*, 242 T*, 247 H
	M1a	$p0$=0.711, $p1$=0.288, $\omega0$=0.109, $\omega1$=1.000	0.366	-6324.428		Not allowed
	M2a	$p0$=0.711, $p1$=0.230, $p2$=0.057, $\omega0$=0.036, $\omega1$=1.000, $\omega2$=1.000	0.366	-6324.428		
	M7	p=0.426, q=1.003	0.297	-6319.041		Not allowed
	M8	$p0$=0.856, p=0.691, q=2.934, $p1$=0.143, ω=1.047	0.310	-6315.448	P < 0.05 (0.0275)	24 L, 26 L, 34 S, 36 T, 52 R, 55 E, 56 F, 67 C*, 78 A, 92 Y, 116 L, 124 V, 134 A, 137 Y, 138 H, 140 P, 184 H, 185 R*, 187 V, 189 C, 203 R, 204 C, 206 K, 209 V, 210 V, 211 R*, 212 L, 213 Q, 214 H*, 215 L, 242 T*, 247 H

(Continued)

Protein	Models[1] Site models	Parameter estimates	ω ratio	ln*L*	LRT[2]	Positively selected (amino acids) sites[3]
HC-Pro	M0	ω=0.085	0.085	-6891.625		None
	M3	*p0*= 0.695, *p1*=0.286, *p2*=0.017 ω*0*=0.010, ω*1*=0.237, ω*2*=0.829	0.090	-6829.734	P < 0.01 (0.0)	Not allowed
	M1a	*p0*=0.943, *p1*=0.056, ω*0*=0.058, ω*1*=1.000	0.1100	-6855.761		Not allowed
	M2a	*p0*=0.943, *p1*=0.044, *p2*=0.011, ω*0*=0.058, ω*1*=1.000, ω*2*=1.000	0.1110	-6855.761		25 Q, 108 I
	M7	*p*=0.252, *q*=2.429	0.089	-6830.501		Not allowed
	M8	*p0*=1.000, *p*=0.252, *q*=2.430, *p1*=0.000, ω=4.608	0.089	-6830.501		4 D, 25 Q, 107 T, 108 I, 245 N
P3	M0	ω=0.153	0.153	-5691.115		None
	M3	*p0*=0.531, *p1*=0.365, *p2*=0.103, ω*0*=0.006, ω*1*=0.161, ω*2*=1.179	0.184	-5559.300	P < 0.01 (0.0)	2 I, 51 Y, 77 R**, 131 H*, 160 P**, 163 N*, 164 T, 176 N**, 179 N**, 190 K, 191 L*, 192 S, 193 A, 194 T*, 195 W, 196 Y**, 197 S*, 198 Y**, 199 R**, 200 A, 202 R**, 203 S**, 204 I*, 206 R**, 207 Y, 212 G, 230 R, 231 G**, 232 A**, 233 Q*, 234 V**, 237 G, 243 S*, 299 Q, 361 D*
	M1a	*p0*=0.864, *p1*=0.135, ω*0*=0.053, ω*1*=1.000	0.181	-5563.445		Not allowed
	M2a	*p0*=0.865, *p1*=0.128, *p2*=0.006 ω*0*=0.054, ω*1*=1.000, ω*2*=2.545	0.192	-5562.836	P > 0.05 (0.543)	179 N, 196 Y, 198 Y
	M7	*p*=0.153, *q*=0.737	0.171	-5567.421		Not allowed
	M8	*p0*=0.904, *p*=0.420, *q*=4.986, *p1*=0.095, ω=1.222	0.184	-5559.396	P < 0.001 (0.0003)	51 Y, 77 R**, 131 H*, 160 P**, 163 N*, 164 T, 176 N**, 179 N**, 190 K, 191 L*, 193 A, 194 T*, 195 W, 196 Y**, 197 S*, 198 Y**, 199 R**, 200 A, 202 R**, 203 S**, 204 I*, 206 R**, 207 Y, 230 R, 231 G**, 232 A**, 234 V**, 243 S*, 299 Q, 361 D*
6K1	M0	ω=0.087	0.087	-704.608		None
	M3	*p0*=0.423, *p1*=0.339, *p2*=0.237, ω*0*=0.000, ω*1*=0.156, ω*2*=1.156	0.090	-703.224	P > 0.05 (0.597)	None
	M1a	*p0*=1.000, *p1*=0.000, ω*0*=0.087, ω*1*=1.000	0.087	-704.608		Not allowed
	M2a	*p0*=1.000, *p1*=0.000, *p2*=0.000 ω*0*=0.087, ω*1*=1.000, ω*2*=12.816	0.087	-704.608		None
	M7	*p*=0.813, *q*=7.883	0.091	-703.647		Not allowed
	M8	*p0*=1.000, *p*=0.813, *q*=7.883, *p1*=0.000, ω=5.067	0.091	-703.647		None
CI	M0	ω=0.050	0.050	-8548.965		None
	M3	*p0*=0.845, *p1*=0.150, *p2*=0.003, ω*0*=0.014, ω*1*=0.238, ω*2*=1.446	0.053	-8488.735	P < 0.01 (0.0)	15 T**, 381 I

(Continued)

Protein	Models[1] Site models	Parameter estimates	ω ratio	ln*L*	LRT[2]	Positively selected (amino acids) sites[3]
CI	M1a	$p0$=0.965, $p1$=0.034, $\omega0$=0.033, $\omega1$=1.000	0.066	-8508.226		Not allowed
	M2a	$p0$=0.965, $p1$=0.032, $p2$=0.001, $\omega0$=0.033, $\omega1$=1.000, $\omega2$=1.000	0.066	-8508.226		None
	M7	p=0.183, q=2.945	0.053	-8493.151		Not allowed
	M8	$p0$=0.995, p=0.222, q=4.067 $p1$=0.004, ω=1.425	0.053	-8489.164	P < 0.05 (0.018)	15 T*, 381 I
6K2	M0	ω=0.042	0.042	-682.505		None
	M3	$p0$=0.197, $p1$=0.738, $p2$=0.064, $\omega0$=0.000, $\omega1$=0.031, $\omega2$=0.371	0.047	-679.229	P < 0.01 (0.0)	None
	M1a	$p0$=0.973, $p1$=0.026, $\omega0$=0.031, $\omega1$=1.000	0.056	-679.986		Not allowed
	M2a	$p0$=0.973, $p1$=0.026, $p2$=0.000 $\omega0$=0.031, $\omega1$=1.000, $\omega2$=20.775	0.056	-679.986		None
	M7	p=0.026, q=9.714	0.048	-679.506		Not allowed
	M8	$p0$=1.000, p=0.260, q=4.714, $p1$=0.000, ω=3.291	0.048	-679.506		None
VPg	M0	ω=0.078	0.078	-2565.512		None
	M3	$p0$=0.862, $p1$=0.137, $p2$=0.000 $\omega0$=0.020, $\omega1$=0.491, $\omega2$=27.610	0.084	-2528.559	P < 0.01 (0.0)	None
	M1a	$p0$=0.919, $p1$=0.080, $\omega0$=0.035, $\omega1$=1.000	0.113	-2534.188		Not allowed
	M2a	$p0$=0.919, $p1$=0.080, $p2$=0.000 $\omega0$=0.035, $\omega1$=1.000, $\omega2$=13.477	0.113	-2534.188		123 T, 164 L
	M7	p=0.142, q=1.409	0.087	-2528.759		Not allowed
	M8	$p0$=0.982, p=0.186, q=2.448 $p1$=0.017, ω=1.341	0.087	-2527.488	P > 0.05 (0.280)	119 G, 123 T, 164 L
NIa	M0	ω=0.087	0.087	-3965.610		
	M3	$p0$=0.538, $p1$=0.414, $p2$=0.046 $\omega0$=0.000, $\omega1$=0.150, $\omega2$=0.789	0.099	-3917.779	P < 0.01 (0.0)	
	M1a	$p0$=0.936, $p1$=0.063, $\omega0$=0.057, $\omega1$=1.000	0.117	-3929.965		Not allowed
	M2a	$p0$=0.936, $p1$=0.065, $p2$= 0.000 $\omega0$=0.057, $\omega1$=1.000, $\omega2$=42.483	0.117	-3929.965		
	M7	p=0.246, q=2.106	0.100	-3919.532		Not allowed
	M8	$p0$=1.000, p=0.246, q=2.106 $p1$=0.000, ω=5.062	0.100	-3919.533		21 V, 104 V, 173 V
NIb	M0	ω=0.079	0.079	-7900.021		None
	M3	$p0$=0.784, $p1$=0.165, $p2$=0.049 $\omega0$=0.019, $\omega1$=0.237, $\omega2$=4.640	0.087	-7815.714	P < 0.01 (0.0)	

(Continued)

Protein	Models[1] Site models	Parameter estimates	ω ratio	lnL	LRT[2]	Positively selected (amino acids) sites[3]
NIb	M1a	*p0*=0.940, *p1*=0.059, *ω0*=0.049, *ω1*=1.000	0.106	-7834.727		Not allowed
	M2a	*p0*=0.940, *p1*=0.038, *p2*=0.021 *ω0*=0.490, *ω1*=1.000, *ω2*=1.000	0.106	-7834.727		98 I, 101 L, 157 L, 266 I, 511 C
	M7	*p*=0.219, *q*=2.144	0.088	-7818.351		Not allowed
	M8	*p0*=1.000, *p*=0.219, *q*=2.148 *p1*=0.000, *ω*=4.620	0.088	-7818.352	P > 0.01 (0.999)	98 I, 100 Y, 101 L, 157 L, 266 I, 290 I, 511 C
CP	M0	*ω*=0.165	0.165	-3311.127		None
	M3	*p0*=0.879, *p1*=0.095, *p2*=0.024 *ω0*=0.061, *ω1*=0.794, *ω2*=2.095	0.181	-3249.592	P < 0.01 (0.0)	24 P, 25 N, 187 V, 193 G*
	M1a	*p0*=0.889, *p1*=0.110, *ω0*=0.061, *ω1*=1.000	0.165	-3251.333		Not allowed
	M2a	*p0*=0.895, *p1*=0.088, *p2*=0.016 *ω0*=0.066, *ω1*=1.000, *ω2*=2.309	0.185	-3249.782	P > 0.05 (0.212)	24 P, 25 N, 187 V, 193 G
	M7	*p*=0.166, *q*=1.795	0.171	-3258.200		Not allowed
	M8	*p0*=0.920, *p*=1.360, *q*=15.048 *p1*=0.079, *ω*=1.369	0.180	-3250.429	P > 0.0001 (0.0004)	9 G, 10 S, 11 T, 24 P**, 25 N**, 26 L*, 58 K, 63 T*, 68 G, 99 L*, 128 I, 138 D**, 187 V**, 193 G**, 217 H, 230 S

[1]Model descriptions according to Yang et al. (2000) M0 (one ration); M3 (discrete); M7 (β); M8 (β plus ω), Wong et al. (2004), Yang et al. (2005) M1a (nearly neutral); M2a (positive selection).

[2]LRTs of M3 vs. M0 are tests of heterogeneity of selection pressures among codon sites. M2a vs. M1a and M8 vs. M7 are tests of positive selection; all assess LRT statistics (2dlnL) against a chi-square distribution with the degrees of freedom equal to the difference in the number of parameters between the nested models being compared.

[3]Amino acid (codon) sites with higher posterior probabilities of P > 95.0 (*) and P > 99.0 (**) undergoing positive selection are shown. Identification of positively selected amino acids is based on either the Naive Empirical Bayes (NEB) approach (under M3) or the Bayes Empirical Bayes (BEB) approach (with the M2a, M8, and branch-site model A).

N lineage; however, some of the motifs including PSYNT (amino acids 160 to 164), KLSATWYSYRAKRSITRY (amino acids 190 to 207) and RGAQV (amino acids 230 to 234) were conserved among the three lineages (Table 3). Except PSYNT motif, the two other ones (amino acids 190 to 207 and amino acids 230 to 234) belong also to the overlapping PIPO protein.

Discussion

PVY is considered one of the most harmful potato viruses. Dependent upon strains, time of infection and cultivars involved the virus can cause serious yield losses reaching to 80% (De Bokx and Huttinga, 1981). Although PVY isolates belonging to PVY^N strain usually cause mild or no obvious symptoms in potato foliage, yield losses may be increased by emergence of new recombinant strains such as PVY^{NTN} which are considered more aggressive than those of PVY^N inducing potato tuber necrotic ring spot disease (PTNRD) (Tian et al., 2011). Furthermore the PVY^{N-Wi} differs in virulence from older PVY isolates; however, serologically it is related to PVY^O isolates (Chrzanowska, 1991). Analysis of previously surveyed potato cultivation areas (Pourrahim et al., 2007) and newly collected isolates during 2009 to 2013 revealed an increase in incidence of PVY^N (unpublished data) in some provinces located in the West and Central parts of Iran. Furthermore, recently Hosseini et al. (2011) showed an increase in the incidence of different strains of PVY including the recombinant isolates from the Southern parts of Iran.

In this srudy, we analyzed the full-length sequences of four N strain isolates from Western and Central parts of Iran (Table S1). The genomes of the Iranian isolates were 9703 to 9707 nucleotides long with the presence of all the reported potyviral motifs. The coding region (polyprotein) of Iranian isolates excluded stop codon consists of 9183

nucleotide. Recombination plays a major role in the evolution of viruses by generating genetic variation, reducing mutational load and producing new viruses (Froissart et al., 2005; Garcia-Arenal and Palukaitis, 2008). The Iranian recombinant PVY^{NTN} isolates are more closely related to the European than North American NTN isolates (Fig. 2 and Fig. 3). Like most of the European isolates (Ogawa et al., 2008), four Iranian NTN isolates analyzed in this study contained three predictable recombination breakpoints between the P1 and HC-Pro, 6K2 and CP, and in the CP cistrons (Table 1) with parental-like isolates belonging to O and N lineages (Fig. 2A and B).

Most of seed potatoes in Iran originated from seed tubers imported from Europe and this suggests the possibility that the Iranian NTN strain originated potentially from an ancestral European PVY strains. Furthermore the spread of PVY mostly occurs from infected to healthy plants within the same potato field via insect vectors (Broadbent et al., 1956, Thresh, 1976). Therefore, occurrence of additional infections in the Iranian potato fields derived from the original infected seed tubers cannot be ruled out (Pourrahim et al., 2007).

Phylogenetic trees showed that four Iranian PVY isolates were clustered in N-Europe subpopulation from N lineage (Fig. 3). The within-population diversity was lower than the between population diversity suggesting the contribution of a recent expansion after a bottleneck, namely a 'founder effect' on diversification of PVY isolates (Table 2, Table 3). Furthermore, nucleotide diversity was higher in C lineage as compared to O and N lineages, which could be explained on the basis of genetic differentiation between three populations. Phylogenetic analyses showed that N-Europe population with closely related isolates appeared in N lineage, as supported by the lowest nucleotide diversity, low pairwise nucleotide diversity and high haplotype diversity (Table 2). A combination for high haplotype diversity and low genetic diversity, assessed by mitochondrial DNA markers, is taken as evidence of a recent population expansion after a genetic bottleneck and this also was found for *Turnip mosaic virus* (TuMV) (Tomitaka and Ohshima, 2006) and *Cauliflower mosaic virus* (CaMV) (Farzadfar et al., 2014). These results are also confirmed by the genome-wide pairwise diversity analysis of the PVY isolates (Fig. 1).

Using PAML analysis most of the sites under positive selection pressure were detected in N lineage compared to those of C and O lineages (Table 4) as previously indicated by Moury and Simon, (2011). This may show the potential of N lineage for better fitness during interactions with hosts, vectors or occupation of new niches as discussed by Karasev et al. (2011). Also, selective pressure on the newly generated recombinant genomes may be so high that only a few recombinant types could survive and succeed in genome propagation (Hu et al., 2009; Karasev et al., 2011).

Regarding the diversity of functions and interactions of different proteins (Moury and Simon, 2011), it is not surprising that evolutionary constraints vary between potyviral proteins. The largest and lowest dn/ds ratios were found for P1 and CI cistrons, respectively (Table 4). Furthermore PAML analysis showed different sites within P1, P3 and CP proteins were under positive selection pressure. Amongst all of the potyvirid proteins, P1 proteinase is the least conserved protein in sequence and the most variable in size (Adams et al., 2005). Mostly this non-conservative protein is thought to have a contribution in successful adaptation of the potyviruses into a wide range of host species (Rohožkoá and Navrátil, 2011; Valli et al., 2007). Also, P1 is a serine protease that self-cleaves at its C-terminus and acts as an accessory factor for genome amplification (Verchot and Carrington, 1995). Based on analysis and observations, Ramírez-Rodríguez et al. (2009) suggested that the determinants of *Tobacco necrosis virus* may be localized at the 3' end of the P1 gene. Most of the amino acid sites under positive selection pressure were found in C-terminal of the P1 which may contribute in the ability of HC-Pro to suppress RNA silencing (Valli et al., 2006) to increase the pathogenicity of heterologous plant viruses in synergistic interactions (Pruss et al., 1997).

Using M3, M2a and M8, BEB predicted the 35 amino acids residue under positive selection pressure for P3 and interestingly most of thse sites overlapped P3N-PIPO protein (Table 4). Studies on P3N-PIPO have indicated that both P3 and P3N-PIPO are essential proteins for potyvirus infection with the potential for involvement in virulence of viruses to overcome resistance mediated by *cyv1* and *sbm-2* genotypes of pea (Choi et al., 2013). In addition, mutations that knock out expression of the PIPO protein in *Turnip mosaic* potyvirus but leave the polyprotein amino acid sequence unaltered are lethal to the virus (Chung et al., 2008; Vijayapalani et al., 2012). However, mutational analysis of the putative *pipo* of *Soybean mosaic virus* suggests disruption of PIPO protein impedes movement, but does not abolish replication in a single cell (Wen and Hajimorad, 2010).

The detailed evolutionary knowledge of eIF4E and VPg makes this coevolutionary pathosystem among the best studied in all of plant virology (Moury et al., 2004). It was shown the VPg of PVY to be under positive selection pressure (Table 4). Thus, mutations at the amino acid positions subjected to positive selection pressure in the VPg of PVY can be expected to play a role in PVY-plant interactions.

Using infectious cDNA molecules from two PVY isolates differing in their virulence, Moury et al. (2004) showed that a single nucleotide change corresponding to an amino acid substitution (Arg119His) in the central part of the viral VPg was involved in virulence toward the *pot-1* resistance. Also, positively selected codons of VPg are directly involved in overcoming resistance eIF4E alleles (Ayme et al., 2007; Charron et al., 2008; Moury et al., 2004).

In PAML analysis, comparison of M8 with M7 for CP showed that M8 fitted the data better than M7 and 16 sites were under positive selection pressure (Table 4). Most of the positive selection sites identified in N-terminal of CP were the same as those previously reported by Moury et al. (2002). The N-terminal part of CP is a notable example of the multifunctionality of potyviral proteins. This part of CP is predicted to be exposed on the virion surface and involves in binding ligands and aphid transmission (Atreya et al., 1995), cell-to-cell and long-distance movement (Dolja et al., 1994, Dolja et al., 1995; Rojas et al., 1997) and/or assembly (Hofius et al., 2007); hence, it can be a target for selection by both host plants and insect vectors.

The first experimental validations of predicted positively selected codon positions were measured for CP of PVY using the impact on virus fitness of nonsynonymous substitutions at codon positions 25 and 68. It was shown that these positive sites were involved in adaptive trade-offs (Moury and Simon, 2011). Also, deletion of 25 amino acids of the CP of *Tobacco etch virus* (TEV) (8 of these amino acids correspond to amino acid positions subjected to positive selection in the PVY genome) reduced the speed of cell to cell movement in tobacco and completely abolished systemic movement (Dolja et al., 1994). Although some an aspartic acid–alanine–glycine (DAG) amino acid triplet, to be highly conserved among potyviruses and essential in their transmission by aphids (Atreya et al., 1995) was under sever negative selection (Moury and Simon, 2011) as indicated in our results.

Our analysis reported in this paper, to the best of our knowledge, provides the first demonstration of population structure in PVY strain N in mid-Eurasia Iran using complete genome sequences that highlights the importance of recombination and selection pressure in the evolution of PVY. The accurate pictures of phylogenetic relationships and comparisons among distantly related virus genomes present an important insight into basic evolutionary mechanisms. Furthermore the appearance of new genetic types not only as a result of recombination, but also mutation, demonstrate a possible high risk that must be taken into account while considering planning for efficient control programs.

Acknowledgments

The authors gratefully acknowledge to Prof. Mohammad Reza Hajimorad (Entomology and Plant Pathology Department, The University of Tennessee, USA) for his critical review and editing of the manuscript. We thank anonymous reviewers for discussion and helpful comments.

References

Adams, M. J., Antoniw, J. F. and Beaudoin, F. 2005. Overview and analysis of the polyprotein cleavage sites in the family *Potyviridae*. *Mol. Plant. Pathol*. 6:471–487.

Atreya, P. L., Lopez-Moya, J. J., Chu, M., Atreya, C. D. and Pirone, T. P. 1995. Mutational analysis of the coat protein N-terminal amino acids involved in *Potyvirus* transmission by aphids. *J. Gen. Virol.* 76:265–270.

Ayme, V., Petit-Pierre, J., Souche, S., Palloix, A. and Moury, B. 2007. Molecular dissection of the *Potato virus Y* VPg virulence factor reveals complex adaptations to the pvr2 resistance allelic series in pepper. *J. Gen. Virol.* 88:1594–1601.

Blanc, S., Lopez-Moya, J. J., Wang, R., Garcia-Lampasona, S., Thornbury, D. W. and Pirone, T. P. 1997. A specific interaction between coat protein and helper component correlates with aphid transmission of a *Potyvirus*. *Virology* 231:141–147.

Broadbent, L., Burt, P. E. and Heathcote, G. D. 1956. The control of potato virus diseases by insecticides. *Ann. Appl. Biol.* 44: 256–273.

Charron, C., Nicola, M., Gallois, J. L., Robaglia, C., Moury, B., Palloix, A. and Caranta, C. 2008. Natural variation and functional analyses provide evidence for coevolution between plant eIF4E and potyviral VPg. *Plant J.* 54:56–68.

Choi, S. H., Hagiwara-Komoda, Y., Nakahara, K. S., Atsumi, G., Shimada, R., Hisa, Y., Naito, S. and Uyeda, I. 2013. Quantitative and qualitative involvement of P3N-PIPO in overcoming recessive resistance against *Clover yellow vein virus* in pea carrying the *cyv1* gene. *J. Virol.* 87:7326–7337.

Chrzanowska, M. 1991. New isolates of the necrotic strain of *Potato virus Y* (PVY^N) found recently in Poland. *Potato Res.* 34: 79–182.

Chung, B. Y.-W., Miller, W. A., Atkins, J. F. and Firth, A. E. 2008. An overlapping essential gene in the *Potyviridae*. *Proc. Nation. Acad. Sci. USA* 105:5897–5902.

De Bokx, J. A. and Huttinga, H. 1981. *Potato virus Y*: description of plant viruses. Kew, England: Commonwealth Mycology Institute. /Association of Applied Biology.

Dolja, V. V., Haldeman, R., Robertson, N. L., Dougherty, W. G. and Carrington, J. C. 1994. Distinct functions of capsid protein in assembly and movement of tobacco etch potyvirus in plants. *EMBO J.* 13:1482–1491.

Dolja, V. V., Haldeman-Cahill, R., Montgomery, A. E., Vanden-

bosch, K. A. and Carrington, J. C. 1995. Capsid protein determinants involved in cell-to-cell and long distance movement of tobacco etch potyvirus. *Virology* 206:1007–1016.

Edwardson, J. R. and Christie, R. G. 1997. Potyviruses. In: Florida Agricultural Experiment Station Monograph Series 18-II-Viruses infecting pepper and other solanaceous crops. Gainesville (FL): University of Florida. p. 424–524.

FAO. 2013. FAOSTAT Database results from FAO website. Food and Agriculture Organization of the United Nations.

Farzadfar, S., Pourrahim, R. and Ebrahimi, H. 2014. A phylogeographical study of the *Cauliflower mosaic virus* population in mid-Eurasia Iran using complete genome analysis. *Arch. Virol.* 159:1329–1340.

Froissart, R., Roze, D., Uzest, M., Galibert, L. and Blanc, S. 2005. Recombination every day: abundant recombination in a virus during a single multi-cellular host infection. *PLoS Biol.* 3:389–395.

Garcia-Arenal, F. and Palukaitis, F. 2008. *Cucumber Mosaic Virus*. In: Mahy B.W.J. and van Regenmortel M.H.V. Encyclopedia of VIROLOGY, THIRD EDITION. Academic Press is an imprint of Elsevier, p. 614–619.

Gibbs, M. J., Armstrong, J. S. and Gibbs, A. J. 2000. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* 16:573–582.

Hall, T. A. 1999. BIOEDIT: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nuc. Acids Sym. Seri.* 41:95–98.

Hofius, D., Maier, A. T., Dietrich, C., Jungkunz, I., Bornke, F., Maiss, E. and Sonnewald, U. 2007. Capsid protein-mediated recruitment of host DnaJ-like proteins is required for *Potato virus Y* infection in tobacco plants. *J. Virol.* 81:11870–11880.

Hosseini, A., Massumi, M., Heydarnejad, J., Hosseini Pour, A. and Varsani, A. 2011. Characterisation of *Potato virus Y* isolates from Iran. *Virus Genes* 42:128–140.

Hu, X., Karasev, A. V., Brown, C. J. and Lorenzen, J. H. 2009. Sequence characteristics of *Potato virus Y* recombinants. *J. Gen. Virol.* 90:3033–3041.

Karasev, A. V., Hu, X., Brown, C. J., Kerlan, C., Nikolaeva, O. V., Crosslin, J. M. and Gray, S. M. 2011. Genetic diversity of the ordinary strain of *Potato virus Y* (PVY) and origin of recombinant PVY strains. *Phytopathology* 101:778–785.

Lorenzen, J. H., Meacham, T., Berger, P. H., Shiel, P. J., Crosslin, J. M., Hamm, P. B. and Kopp, H. 2006. Whole genome characterization of *Potato virus Y* isolates collected in the western USA and their comparison to isolates from Europe and Canada. *Arch. Virol.* 151:1055–1074.

Martin, D. P., Lemey, P., Lott, M., Moulton, V. and Posada, D. 2010. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* 26:2462–2463.

Moury, B. and Simon, V. 2011. dN/dS-based methods detect positive selection linked to trade-offs between different fitness traits in the coat protein of *Potato virus Y*. *Mol. Biol. Evol.* 28:2707–2717.

Moury, B., Morel, C., Johansen, E. and Jacquemond, M. 2002. Evidence for diversifying selection in *Potato virus Y* and in the coat protein of other potyviruses. *J. Gen. Virol.* 83:2563–2573.

Moury, B., Morel, C., Johansen, E., Guilbaud, L., Souche, S., Ayme, V., Caranta, C., Palloix, A. and Jacquemond, M. 2004. Mutations in *Potato virus Y* genome-linked protein determine virulence toward recessive resistances in *Capsicum annuum* and *Lycopersicon hirsutum*. *Mol. Plant-Microbe Interact.* 17: 322–329.

Ogawa, T., Tomitaka, Y., Nakagawa, A. and Ohshima, K. 2008. Genetic structure of a population of *Potato virus Y* inducing potato tuber necrotic ringspot disease in Japan; comparison with North American and European populations. *Virus Res.* 131:199–212.

Pourrahim, R., Farzadfar, Sh., Golnaraghi, A. R. and Ahoonmanesh, A. 2007. Incidence and distribution of important viral pathogens in some Iranian potato fields. *Plant Dis.* 91:609–615.

Pruss, G., Ge, X., Shi, X. M., Carrington, J. C. and Vance, V. B. 1997. Plant viral synergism: the potyviral genome encodes a broad-range pathogenicity enhancer that transactivates replication of heterologous viruses. *Plant Cell* 9:859–868.

Ramírez-Rodríguez, V. R., Aviña-Padilla, K., Frías-Treviño, G., Silva-Rosales, L. and Martínez-Soriano, P. 2006. Presence of necrotic strains of *Potato virus Y* in Mexican. *Virol. J.* 48:1–7.

Rohožkoá, J. and Navrátil, M. 2011. P1 peptidase a mysterious protein of family *Potyviridae*. *J. Biosci.* 36:189–200.

Rojas, M. R., Zerbini, F. M., Allison, R. F., Gilbertson, R. L. and Lucas, W. J. 1997. Capsid protein and helper component-proteinase function as *Potyvirus* cell-to-cell movement proteins. *Virology* 237:283–295.

Rozas, J., Sanchez-DeIBarrio, J. C., Messeguer, X. and Rozas, R. 2003. dnasp, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496–2497.

Shukla, D., Ward, C. W. and Brunt, A. A. 1994. The *Potyviridae*, CAB International, Wallingford.

Singh, R. P., Valkonen, J. P. T., Gray, S. M., Boonham, N., Jones, R. A. C., Kerlan, C. and Schubert, J. 2008. Discussion paper: The naming of *Potato virus Y strains infecting potato*. *Arch. Virol.* 153:1–13.

Singh, R. P. 1992. Incidence of the tobacco veinal necrotic strain of *Potato virus Y* (PVY^N) in Canada in 1990 and 1991 and scientific basis for eradication of the disease. *Canadian Plant Dis. Sur.* 72:113–119.

Tamura, K., Dudley, J., Nei, M. and Kumar, S. 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24:1596–1599.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28: 2731–2739.

Thresh, J. M. 1976. Gradiants of plant virus diseases. *Ann. Appl. Biol.* 82:381–406.

Tian, Y. P., Liu, J. L., Zhang, C. L., Liu, Y. Y., Wang, B., Li, X. D., Guo, Z. K. and Valkonen, J. P. T. 2011. Genetic diversity of *Potato virus Y* infecting tobacco crops in China. *Phytopathology* 101:377–387.

Tomitaka, Y. and Ohshima, K. 2006. Phylogeographical study of the *Turnip mosaic virus* population in East Asia reveals an 'emergent' lineage in Japan. *Mol. Ecol.* 5:4437–4457.

Urcuqui-Inchima, S., Haenni, A. L. and Bernardi, F. 2001. *Potyvirus* proteins: a wealth of functions. *Virus Res.* 74:157–175.

Valli, A., Lopez-Moya, J. J. and Garcia, 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., Martín-Hernández, A. M., López-Moya, J. J. and García, J. A. 2006. RNA silencing suppression by a second copy of the P1 serine protease of *Cucumber vein yellowing ipomovirus* (CVYV), a member of the family *Potyviridae* that lacks the cysteine protease HCPro. *J. Virol.* 80:10055–10063.

Verchot, J. and Carrington, J. C. 1995. Evidence that the *Potyvirus* P1 proteinase functions in trans as an accessory factor for genome amplification. *J. Virol.* 69:3668–3674.

Vijayapalani, P., Maeshima, M., Nagasaki-Takekuchi, N. and Miller, W. A. 2012. Interaction of the trans-frame *Potyvirus* protein P3NPIPO with host protein PCaP1 facilitates *Potyvirus* movement. *PLoS Path.* 8:e1002639.

Weiller, G. F. 1998. Phylogenetic profiles: a graphical method for detecting genetic recombinations in homologous sequences. *Mol. Biol. Evol.* 15:326–335.

Wen, R. H. and Hajimorad, M. R. 2010. Mutational analysis of the putative *pipo* of *Soybean mosaic virus* suggests disruption of PIPO protein impedes movement. *Virology* 400:107.

Wong, W. S. W., Yang, Z., Goldman, N. and Nielsen, R. 2004. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics* 168:1041–1051.

Yang, Z. 2007. PAML4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24:1586–1591.

Yang, Z., Nielsen, R., Goldman, N. and Pedersen, A. K. 2000. Codon substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155:431–449.

Yang, Z., Wong, W. S. W. and Nielsen, R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* 22:1107–1118.