Acta virologica **61**: 161 – 174, 2017 doi:10.4149/av_2017_02_05

Comparative genetic diversity of potato virus Y populations based on coat protein gene

H. HOSSEINI[1], M. MEHRVAR[1], M. ZAKIAGHL[1*], M. SIAMPOUR[2]

[1]Department of Plant Pathology, College of Agriculture, Ferdowsi University of Mashhad, Iran; [2]Department of Plant Protection, College of Agriculture, Shahrekord University, Iran

January 25, 2016; revised September 5, 2016; accepted March 15, 2017

Summary. – Potato virus Y (PVY) is an important plant pathogen with a wide host range including economically important crops of potato, tobacco, tomato, and pepper. The coat protein gene has been commonly used in studying molecular biology of plant viruses including PVY. In this study, we used a large dataset of CP sequences from isolates collected across the world to assess the detailed molecular evolution of PVY populations with a focus on the Iranian PVY population. Phylogenetic analysis showed that the world PVY population had two major lineages (O:C and N:NTN); each comprising several divergent sublineages. Results showed that the Iranian PVY isolates were distributed across the tree suggesting polyphyletic origin of the Iranian PVY population. Statistical analysis revealed great genetic differences between pairs of the PVY phylogenetic populations. Host populations and also geographical populations of PVY were genetically differentiated. The extent of the genetic diversification among PVY host and geographical populations were mild or moderate. Purifying selection was detected on the CP gene sequences of the PVY populations, suggesting that most of the mutations in the gene were harmful, thereby were eliminated by natural selection. We also detected a variety of recombination patterns to occur along the CP gene of the PVY strains. A significant number of the Iranian PVY isolates were found to be recombinant. Different analyses suggest that Iranian PVY population is highly diverse. In conclusion, results of this study demonstrated that different factors including mutation, host adaptation, geographical distinction and selection pressure shaped the genetic structure of the PVY populations.

Keywords: potato virus Y; genetic variability; genetic differentiation; evolution

Introduction

Potyviruses are the largest family of plant viruses. They have flexuous filamentous particles containing a single-strand, positive sense RNA molecule of about 9.7 kb (Shukla *et al.*, 1994; Fauquet *et al.*, 2012). Their genome has a viral-encoded VPg protein at its 5' end followed by a single open reading frame (ORF) and a poly-A tail at the 3' end. The viral genome is translated into a single large polyprotein, and then cleaved into at least ten multifunctional proteins by viral proteinases (Boonham *et al.*, 2002; Gibbs and Ohshima 2010; Martin *et al.*, 2010; Fauquet *et al.*, 2012). Moreover, during translation, there is a frame shift in the P3 gene resulting in production of a new short polypeptide (PIPO) (Chung *et al.*, 2008). Host adaptation determines the length of PIPO among potyviruses. (Hillung *et al.*, 2013). Some members of potyviruses have another ORF, named PISPO, which is a suppressor of gene silencing (Clark *et al.*, 2012). However, existence of PISPO in PVY genome is still to be determined.

Potato virus Y (PVY) is the type species of the genus *Potyvirus*. It is responsible for serious diseases in a wide range of plant species, mostly from the family *Solanaceae*, with worldwide distribution. Like other potyviruses, PVY is transmitted by aphids in a non-persistent manner. Because of the wide host range and prevalence of the virus, PVY is probably the most destructive and widespread plant pathogen across the world (Shukla *et al.*, 1994).

*Corresponding author. E-mail: zakiaghl@ferdowsi.um.ac.ir; phone: +98-511-8795620.

Abbreviations: PVY = potato virus Y

To date, several distinct strains of PVY have been identified according to their biological, serological and genome sequence analyses. These strains are designated as PVY^{O} (the ordinary strain), PVY^{N} (the tobacco venial necrosis) and PVY^{C} (the stipple streak strain; Moury *et al.*, 2002). Furthermore, three other strains including of PVY^{NTN}, $PVY^{N\text{-}Wi}$ and $PVY^{N:O}$ have been identified as being generated by recombination between PVY^{O} and PVY^{N} (Le Romancer *et al.*, 1994; Glais *et al.*, 2002). PVY^{NTN} induces potato tuber necrotic ringspot disease (PTNRD). Since the first report from Hungary in 1980s, the PVY^{NTN} recombinant has been frequently identified within the PVY populations worldwide (Le Romancer *et al.*, 1994; Karasev and Gray, 2013). The PVY^{NTN} has two main recombination patterns. The first one (PVY^{NTNa}) contains three recombination junctions (RJs), while the second structure (PVY^{NTNb}) has one RJ in P1 in addition to the three RJs in common with PVY^{NTNa} (Chikh Ali *et al.*, 2010).

Molecular evolutionary studies of viruses have shown the impacts of mutation, recombination, selection pressure and host adaptation in dynamics of viral populations. These studies have shed light on the important features of viral biology such as changes in virulence, geographical spread and adaptation to new hosts or emergence of a new virus epidemic. Comparisons of the genetic structure of different populations of a virus species can determine factors affecting ecology, phylogeny and phylogeographic structure of the populations across the world (Bermingham and Moritz, 1998). Such studies provide useful information for designing better strategies aimed at controlling of the viruses (Jones, 2009; Elena *et al.*, 2011). Although increasing number of studies on the population structure of animal and human RNA viruses is available, the population structure of plant viruses is poorly understood (Garcia-Arenal *et al.*, 2001).

PVY is reported to be one of the viruses responsible for the severe yield losses in Iran annually (Pourrahim *et al.*, 2007). Based on the host range, symptomatology and serological traits, PVY^{O}, PVY^{NW}, PVY^{C}, PVY^{N} and PVY^{NTN} strains have been recognized in potato fields of Iran. Furthermore, the sequence of P1, coat protein (CP) and 3' UTR of some of the Iranian PVY isolates has been reported (Hosseini *et al.*, 2011).

To study the population structure of PVY in Iran, we analyzed the biological properties and CP sequence variability of Iranian PVY isolates (including 11 new sequences provided in this study) in comparison with PVY populations from other countries. We focused on coat protein gene (CP) because of its multifunctional properties, sequence variability and availability of large sequence data for most of the PVY isolates reported from around the world (Ogawa *et al.*, 2008; Visser and Bellstedt, 2009; Moury and Simon, 2011).

To evaluate, in more detail, the molecular evolution of PVY populations, we compared the CP gene sequence of the Iranian PVY isolates with counterparts reported from different parts of Europe, North and South America, Africa and Japan (n = 542).

Materials and Methods

Virus source and serological diagnosis of PVY. A total of 185 potato leaf samples were collected during July 2013 to July 2014 from the potato fields of Khorasan Razavi, Northern Khorasan and Fars provinces of Iran. The potato plants were sampled randomly regardless of their symptoms. Specific polyclonal antibody against PVY (DSMZ, Germany) was used to detect the virus in fresh leaf crude extracts by DAS-ELISA (Clark and Adams, 1977).

Samples, RT-PCR, cloning and sequencing. Total plant RNA was extracted from fresh leaves using Total RNA isolation kit (Denazist Asia-Iran) following the manufacturer's instructions. The First cDNA strand was synthesized using antisense primer (PVY-CPR) and moloney murine leukemia virus (MMuLV) reverse transcriptase (Fermentas, Lithuania). 5 µl of the purified RNA were mixed with reverse transcription mixture (50 mmol/l Tris-HCl pH 8.3, 50 mmol/l KCl, 4 mmol/l $MgCl_2$, 10 mmol/l dithiothritol, 1 mmol/l of each dNTP, 200 U of MMuLV reverse transcriptase) and incubated at 42°C for 1 h. The complete length of the coat protein (CP) gene was amplified using specific primers PVY-CPF (5'-GCTTTCACTGAAATGATGGT-3') and PVY-CPR (5'-GTTTTCCCAGTCACGACTTTTTTTTTT-3') by Taq DNA Polymerase Master Mix Red (Amplicon, Denmark). These primers amplify the complete CP gene of the PVY (~800 bp length; Nie and Singh, 2003). PCR products were analyzed on 1% agarose gel and purified from the gel using the Qiaquick gel extraction kit (Qiagen, Germany). The purified products were ligated into pTG19 vector (Vivantis, Malaysia) according to the manufacturer's protocol. Plasmids were transformed into *Escherichia coli* strain DH5a and the recombinant plasmids were purified from the bacterial cells using Plasmid DNA Isolation Kit (Denazist-Iran). Finally, the purified recombinant plasmids were bi-directionally sequenced using pUC-M13 universal primers at Macrogen Inc. (Seoul, South Korea). Consensus sequences were verified using the BLAST program in NCBI database.

Phylogenetic analysis. Due to space limitation to show the tree in a single page, preliminary neighbor joining clustering of isolates was carried out using SDT v.1 software. Then 150 representatives out of 542 PVY-CP sequences were selected and phylogenetic tree was reconstructed. An alignment of the CP gene sequences of 150 PVY isolates, including 26 isolates from Iran (11 new sequences obtained in this study) and 124 isolates from other countries (Table 1), was generated using the Muscle module in MEGA v.5 (Tamura *et al.*, 2011). Details of the PVY isolates, their country of origin, and GenBank Acc. Nos. are shown in Table 1. MEGA v.5. was used for generating phylogenetic tree by the Maximum likelihood method with HKY+G4 nucleotide substitution model as the best fitted model. Integrity of the evolutionary relationships was assessed by

Table 1. Accession numbers and origins of potato virus Y isolates whose coat protein were used in this study

Populations	Origin	Number	Acc. Nos.
Middle East	Iran	26	123cm,145,156,168,194,211,21,221,52cm,61,EU713856,HM243471,HM243472,HM243473,HM243474,HM243475,HM243476,HM243477,HM243478,HM243479,HM243480,HM243481,HM243482,HM243483,HM243484,IRAN
	Syria	14	AB185831,AB185832,AB185833,AB185833,AB256029,AB270705,AB270705,AB295475,AB295477,AB29547,AB461450,AB461451,AB461452,AB461452
	Iraq	13	JQ026006,JQ026007,JQ026009,JQ026010,JQ026011,JQ026012,JQ026013,JQ026014,JQ026015,JQ026016,JQ026017,JQ026018,JQ026019
	Jordan	6	EU073854,EU073855,EU073856,EU073857,EU073858,EU073859
	India	2	AF118153,JN034046
	Pakistan	1	JQ518267
Far East	Japan	45	AB331515,AB331516,AB331517,,AB331518,AB331519,AB331538,AB331539,AB331540,AB331541,AB331542,AB331543,AB331544,AB331545,AB331546,AB331547,AB331548,AB331549,AB331550,AB702945,AB702950,AB702951,AB702952,AB702953,AB702954,AB702955,AB702956,AB711143,AB711144,AB711145,AB711146,AB711147,AB711148,AB711149,AB711150,AB711151,AB711152,AB711153,AB711154,AB711155,AB714134,AB714135,AB719459,D12539,D12570
	China	34	AJ488834,AJ488834,AM931254,EF592514,EF592514,EF592515,EF592515,EF592516,EF592516,EF592521,EF592521,EF592525,EF592525,EF592526,EF592526,EU182576,EU719648,EU719650,FJ423031,FJ423032,GQ200836,FJ766535,HM590405,HM590406,HM590407,HQ603083,HQ631374,JQ663997,JQ673517,JX872404,JX872405,KJ634024,KJ801915,PVU25672
	Vietnam	8	AM411502,AM411503,AM411504,DQ925435,DQ925437,FM200035,FM201468,FM201468
	South Korea	1	EU885418
Africa	Egypt	2	AF522296,AF52229
	South Africa	88	GQ853593,GQ853594,GQ853595,GQ853596,GQ853597,GQ853598,GQ853599,GQ853600,GQ853601,GQ853601,GQ853602,GQ853603,GQ853606,GQ853607,GQ853608,GQ853609,GQ853610,GQ853611,GQ853612,GQ853613,GQ853614,GQ853615,GQ853616,GQ853617,GQ853618,GQ853619,GQ853620,GQ853621,GQ853622,GQ853623,GQ853624,GQ853625,GQ853626,GQ853627,GQ853628,GQ853629,GQ853630,GQ853631,GQ853632,GQ853633,GQ853634,GQ853635,GQ853636,GQ853637,GQ853638,GQ853639,GQ853640,GQ853641,GQ853642,GQ853643,GQ853644,GQ853645,GQ853646,GQ853647,GQ853648,GQ853649,GQ853650,GQ853651,GQ853652,GQ853653,GQ853655,GQ853656,GQ853657,GQ853658,GQ853659,GQ853660,GQ853661,GQ853662,GQ853663,GQ853664,GQ853665,GQ853666,GQ853667,JN936418,JN936420,JN936422,JN936423,JN936425,JN936427,JN936429,JN936430,JN936432,JN936435,JN936437,JN936438,JN936439,JN936440,JN936441
Europe	United Kingdom	41	AF325927,AF325928,AJ390285,AJ390285,AJ390286,AJ390286,AJ390287,AJ390287,AJ390288,AJ390289,AJ390290,AJ390291,,AJ390292,AJ390294,AJ390295,AJ390296,AJ390297,AJ390301,AJ390302,AJ390303,AJ390304,AJ390305,AJ390306,AJ390308,AJ390309,AJ585195,AJ585196,AJ585197,AJ585198,EF016294,JQ954337,JQ954338,JQ954339,KC614702,KC634004,KC634005,KC634006,KC634007,KC634008,KC634009
	Germany	29	AJ889867,AJ889868,AJ890347,AJ890350,AM113988,HE608963,HE608964,JQ954295,JQ954296,JQ954297,JQ954298,JQ954299,JQ954300,JQ954301,JQ954302,JQ954303,JQ954304,JQ954305,JQ954307,JQ954308,JQ954309,JQ954310,JQ954311,JQ954312,JQ954313,JQ954314,JQ954384,JQ954387,JQ954342
	Poland	11	AJ889866,AJ890342,EF558545,EF558545,FJ666337,JF795485,JF804780,JF804781,JF804785,JF804786,JF804787
	Switzerland	9	JQ954331,JQ954332,JQ954333,JQ954334,JQ954335,JQ954336,JQ954393,JQ954394,X97895
	Belgium	8	JQ969033,JQ969034,JQ969035,JQ969036,JQ969037,JQ969039,JQ969040,JQ969041
	Czech Republic	8	JQ954343,JQ954344,JQ954345,JQ954346,JQ954347,JQ954348,JQ954340,JQ954341
	Italy	5	JQ954322,JQ954323,JQ954324,JQ954325,JQ954326
	Slovenia	4	AJ390293,AJ585342,JQ954376,JQ954377
	Netherlands	4	JQ954327,JQ954328,JQ954329,JQ954330
	Hungary	3	AJ390300,M95491,M95491
	France	2	AJ890348,HM991454
	Portugal	1	AJ390307
	Denmark	1	AJ390298,AJ390299
	Netherlands	1	EU563512
	Latvia	1	GQ496607
	Greece	1	JQ954321
	Finland	1	JX424837

Table 1. (continue)

Populations	Origin	Number	Acc. Nos.
N. America	USA	108	AJ390309,AY884982,AY884983,AY884984,FJ666337,AY884985,DQ008213,DQ157178,DQ157179,DQ157180,EF026074,EF026075,EF026076,FJ204164,FJ204165,FJ204166,FJ643477,FJ643478,FJ643479,HQ912862,HQ912863,HQ912864,HQ912865,HQ912866,HQ912867,HQ912868,HQ912869,HQ912870,HQ912871,HQ912872,HQ912873,HQ912874,HQ912875,HQ912876,HQ912877,HQ912878,HQ912879,HQ912880,HQ912881,HQ912882,HQ912883,HQ912884,HQ912885,HQ912886,HQ912887,HQ912888,HQ912889,HQ912890,HQ912891,HQ912892,HQ912893,HQ912894,HQ912895,HQ912896,HQ912897,HQ912898,HQ912899,HQ912900,HQ912901,HQ912902,HQ912903,HQ912904,HQ912905,HQ912906,HQ912907,HQ912908,HQ912909,HQ912910,HQ912911,HQ912912,HQ912913,HQ912914,HQ912915,JQ954349,JQ954350,JQ954351,JQ954352,JQ954353,JQ954354,JQ954355,JQ954356,JQ954357,JQ954358,JQ954359,JQ954360,JQ954361,JQ954362,JQ954363,JQ954364,JQ954365,JQ954366,JQ954367,JQ954368,JQ954369,JQ954370,JQ954371,JQ954372,JQ954373,JQ954374,JQ954375,JQ954385,JQ954386,JQ954388,JQ954389,JQ954390,JQ954391,JQ954392,U91747
	Canada	11	AF126258,AY166866,,AY166867,,AY512655,AY745491,AY745492,HM367075,HM367076,PVU09509,U09508,U09509
	Mexico	1	KF850513
S. America	Brazil	11	AF255659,AF255660,AF525081,AY840082,JF928458,JF928459,JF928460,JQ924285,JQ924286,JQ924287,JQ924288
	New Zealand	2	AM268435,DQ217931

1000 bootstrap replicates. Pairwise distance comparisons of the CP gene sequences were computed using the Tamura-Nei model in MEGA v. 5.0.

Population genetics analysis. DnaSP version 5.10.01 (Libardo and Rozas, 2009) was used to estimate the average pairwise nucleotide diversity (π, average distances between pairs of sequences; Tajima 1983), number of polymorphic site (S), total number of mutations (η), average number of nucleotide differences among sequences from the same population (K), haplotype diversity (Hd, number of haplotypes within a sample), and the ratio of non-synonymous to synonymous nucleotide diversity (dN/dS), also known as ω ratio. The nucleotide diversity (π) may range between 0.0 (no variation) and 0.100 (highest variation between sequences). The haplotype diversity may range from 0.0 and 1.000. This program was also used for Tajima's D (Tajima 1989), Fu and Li's D* and F* (Fu and Li, 1993) tests of neutrality. Tajima's *D* test is based on the differences between the numbers of segregating sites and the average number of nucleotide differences. Fu and Li's *D** test is based on the differences between the numbers of mutations that appear in only one sequence and the total numbers of mutations. Fu and Li's *F** test is based on the differences between the numbers of singletons and the average number of nucleotide differences between every pair of sequences.

Tests of population differentiation. Statistical tests of population differentiation including Kst*, Z*, Snn and F_{ST}, were calculated using DnaSP version 5.10.01. The Kst* test statistic of genetic differentiation (Hudson, 2000) is expected to be near zero if there is no genetic differentiation (null hypothesis). But the null hypothesis is rejected, when Kst* is supported by a small P-value (0.05) (Tsompana *et al.*, 2005). The Z statistic is calculated from ranking distances between all pairs of sequences. The average ranks for those from within two populations are summed, and the sum is weighted. The Z* statistic is a logarithmic variant of the Z statistic (Hudson, 2000). Also, small values of Z lead to rejection of the null hypothesis (no genetic differentiation). The nearest neighbor statistic (Snn) was used to evaluate the frequency of the nearest neighbors of sequences found in the same group (Hudson, 2000). Snn statistic may range between 1.0, when populations from distinct groups were genetically different, to 0.5 when a population is panamictic (Tsompana *et al.*, 2005). F_{ST} is the coefficient of gene differentiation or fixation index, which measures the extent of inter-population diversity with values ranging between 0 (indicating no differentiation between the populations) and theoretical maximum of 1 (when the populations are clearly differentiated; Hudson, 2000). However, the observed F_{ST} is much less than 1, even in highly differentiated populations. Statistical significance for all three tests (Z*, Snn and F_{ST}) was established using 1000 permutations test.

Tests of recombination. Alignments of 542 PVY-CP sequences were analyzed for intraspecies recombination events using Recombination Detection Program (RDP v.4.10 beta) with default parameters (highest acceptable probability value = 0.05; Martin *et al.*, 2010). The RDP4 software detects the occurrence of robust recombination events using a suite of methods including Rdp, Geneconv, Bootscan, Maxchi, Chimaera, 3Seq, Siscan, Lard and Phylpro. Bootscan, Rdp and Siscan are phylogenetic methods, Geneconv, Maxchi, Chimaera and Lard are substitution-based methods, and Phylpro is a distance comparison method. Recombination sites detected at least by four methods were considered as "significant recombination events" and those detected by fewer methods were considered as "tentative recombination events" (Heath *et al.*, 2006).

Biological characterization of the PVY isolates. The partial host range of four Iranian PVY isolates were compared by their inoculation to potato (*Solanum tuberosum*), tomato (*Lycopersicum*

esculentum), tobacco (*Nicotiana tabacum* var Turkish, *Nicotiana tabacum* var Xanthi), pepper (*Capsicum annuum*) and *Chenopodium quinoa* seedlings. To this purpose, potato leaf tissues infected by these PVY isolates were separately ground in 5 volumes of 50 mmol/l phosphate buffer at pH 7.5 containing 2% polyvinylpyrrolidone (PVP) and 0.05% 2-mercaptoethanol. The extracts were mechanically inoculated on carborandum-dusted leaves. Two weeks post inoculation plants were examined for symptoms and tested for PVY infection using RT-PCR as described above.

Results

Identification and sequencing of CP genes of PVY isolates

From a total of 185 plants sampled, PVY was detected by DAS-ELISA in 46 samples collected from Khorasan-Razavi, Northern-Khorasan (North-East), and Fars (South-East) provinces of Iran. PVY was detected in potato plants showing symptoms of yellowing, mosaic, leaf crinkling as well as in the asymptomatic plants.

The complete nucleotide sequence of the CP gene from 11 Iranian PVY isolates were determined and deposited in GenBank (Table 2).

Phylogenetic positions of the PVY populations

The maximum likelihood phylogenetic tree of the CP gene from 150 PVY isolates, including 26 isolates from Iran (11 sequences provided in this study and 15 sequences obtained from GenBank) and 124 isolates from other countries (obtained from GenBank), revealed the segregation of PVY isolates into two main evolutionary divergent clades (Fig. 1).

These two major clades, designated as $PVY^{O:C}$ and $PVY^{N:NTN}$ included 84 and 66 isolates, respectively. The O:C clade was partitioned into two sister clades designated as O (PVY^{O}) and C (PVY^{C}).

The PVY^{O} clade (Ordnary clade) was divided into five distinct subclades designated as PVY^{O}-O5, PVY^{O}-OJ1, PVY^{O}-O1, PVY^{O}-O2 and PVY^{O}-O3. The PVY^{O}-O2 and PVY^{O}-O3 were the predominant genotypes within the PVY^{O} clade. Furthermore, the $PVY^{N:NTN}$ clade included two subclades termed as N-Europe and N-North America (Fig. 1). The N-Europe subclade could be further segregated into two groups of PVY^{NTNa} and PVY^{NTNb}.

As shown by the phylogenetic tree, PVY isolates were arranged between the two main clades ($PVY^{O:C}$ and $PVY^{N:NTN}$) irrespective of their country of origin. This suggests that geographical origin has no significant effect on the phylogenetic divergence of PVY isolates into two main clades. The Iranian isolates were also found to be distributed between these two clades. Of 26 isolates from Iran, 15 isolates grouped within the $PVY^{O:C}$ clade and 11 isolates grouped within the $PVY^{N:NTN}$ clade (Fig. 1).

The PVY^{O}-O2 genotype (n = 10) was the most common genotype among Iranian isolates of PVY^{O}. None of the Iranian PVY^{O} isolates placed in PVY^{O}-O5 and PVY^{O}-OJ1 clades. The majority of the Iranian $PVY^{N:NTN}$ genotypes were clustered with PVY^{NTNa} group in N-Europe subclade.

The mean sequence distance among isolates of $PVY^{O:C}$ and $PVY^{N:NTN}$ clades was 3.6% and 3.4%, respectively. The mean sequence distance between these two clades was 9.8%.

The Iranian PVY isolates in the $PVY^{O:C}$ clade shared 96.46±0.52% CP nucleotide sequence identity with each other and 96.01±1.63% with other isolates of this clade. On the other hand, the Iranian PVY isolates in the $PVY^{N:NTN}$ had 93.76±3.39% CP nucleotide sequence identity with each other and 96.13±2.4% with other isolates of the clade.

Population characteristics and polymorphism

To study the diversity of the PVY population, all PVY CP sequences (determined in this study or obtained from GenBank; n = 542), were categorized into 3 population groups

Table 2. Characterization of Iranian potato virus Y isolates reported in this study

Samples	Acc. No.	Host	Strain	DAS-ELISA	RT-PCR
IRAN	LN908250	potato	NTN	+	+
52cm	LN907865	potato	NTN	+	+
123cm	LN907866	potato	NTN	+	+
168	LN907864	potato	NTN	+	+
156	LN907860	potato	N:O	+	+
211	LN907861	potato	SYR-NB/O	+	+
21	LN880858	potato	O	+	+
194	LN907859	potato	O	+	+
221	LN907862	potato	O	+	+
145	LN908252	potato	O	+	+
61	LN908251	potato	O:C	+	+

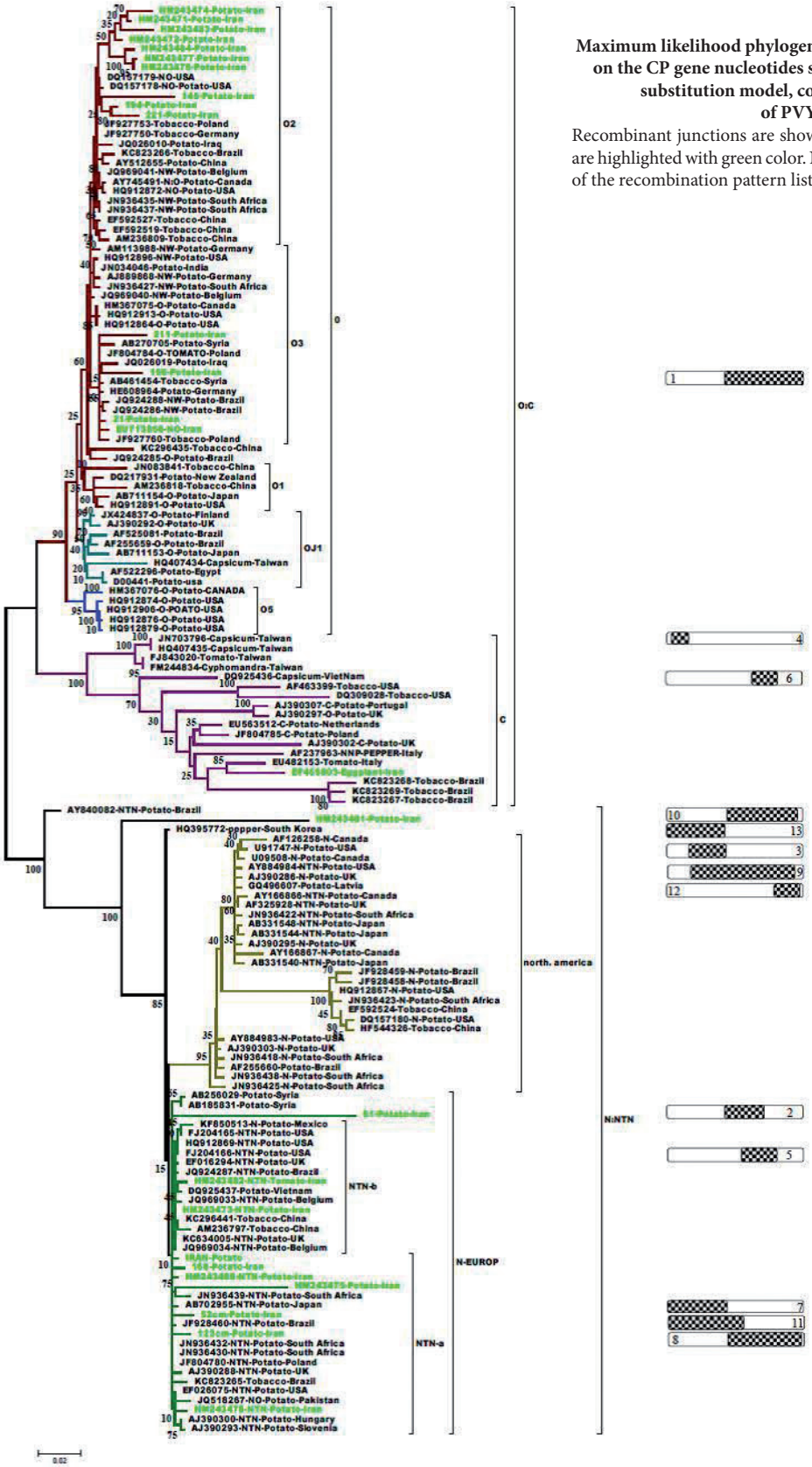

Fig. 1

Maximum likelihood phylogenetic analysis of 150 PVY isolates based on the CP gene nucleotides sequences using HKY+G4 nucleotide substitution model, coupled with schematic diagram of PVY recombinants

Recombinant junctions are shown as checker-board box. Iranian isolates are highlighted with green color. Numbers in rectangular represent number of the recombination pattern listed in Table 7.

based on their phylogenetic relationship (4 phylogenetic populations), host adaptation (3 host populations) and geographical origin (6 geographical populations; Table 3).

Among 4 PVY phylogenetic populations, strains of PVY^{C} had the most average number of CP nucleotide differences (k = 55 nucleotides) and the highest overall nucleotide diversity ($\pi = 0.0691$). The largest number of segregating sites (S = 350) and mutations within the segregating sites ($\eta = 442$) were observed within the CP nucleotide sequences of the PVY^{O} lineage (Table 3). The PVY^{NTN} population had the lowest values of k (10.50) and π (0.0131; Table 3). On the other hand, the ratio of nonsynonymous nucleotide diversity to synonymous nucleotide diversity (ω ratio) was <1 for all phylogenetic populations. The highest and the lowest ratios were obtained for the PVY^{NTN} and PVY^{N} populations with the values of 0.351 and 0.046, respectively. The dN/dS value of the CP gene from different PVY phylogenetic populations ranged between 0.046 and 0.35 (Table 3).

PVY host populations included those isolated from potato, tobacco and other hosts (e.g. tomato; Table 3). Among 3 PVY host populations, the greatest average number of CP nucleotide differences (k = 61) and the greatest overall nucleotide diversity ($\pi = 0.0771$) was observed in the PVY population isolated from other hosts (other than potato and tobacco; Table 3). The CP sequence analysis showed that PVY population of potato had the highest number of segregating sites (S = 442) and mutations ($\eta = 611$; Table 3). The k and π statistics were the lowest for the PVY population of tobacco with values of 45 and 0.0572, respectively (Table 3). The dN/dS value of the CP gene sequence from the three PVY host populations ranged from 0.08 to 0.1 (Table 3).

Whereas the PVY population of tobacco had the lowest ω ratio (0.0836) in the CP gene, the value was highest for PVY strains isolated from hosts other than potato and tomato (other hosts, $\omega = 0.1031$).

We also defined 7 geographical populations based on the geographical origin of the PVY isolates (Table 3). The CP Molecular variability revealed that the South American PVY population had the highest number of nucleotide differences (k = 65) and the highest overall nucleotide diversity ($\pi = 0.0814$). The largest number of segregating sites (S = 250) and mutations within the segregating sites ($\eta = 297$) was observed within the Iranian PVY population (Table 3). Moreover, the dN/dS ratio (ω) was <1 for all of the geographical populations except for African population ($\omega = 1.03$). In this regard, the largest (1.031) and lowest (0.095) ω were observed in PVY populations of Africa and South America, respectively (Table 3).

Genetic differentiation of populations

We evaluated the genetic distinction of PVY populations defined in 3 categories: phylogenetic populations, host populations, and geographical populations. In pursue of this goal the PVY-CP gene sequences were subjected to four independent statistical tests of population differentiation. The null hypothesis (no genetic differentiation) can be rejected if the statistics Kst*, K, Z, and Snn for the majority

Table 3. Genetic characteristics of potato virus Y coat protein from different populations

	Population	No. seq.	S	η	k	π	SS	NS	dS	dN	dN/dS
Phylogenetic population	All	542	496	722	49.5229	0.0621	179.76	609.24	0.2611	0.0207	0.0794
	O	311	**350**	**442**	24.0685	0.0300	182.09	618.91	0.0778	0.0180	**0.2323**
	C	21	190	211	**55.3455**	**0.0691**	**182.24**	618.76	**0.2507**	**0.0284**	0.1133
	N	33	97	103	27.1014	0.0339	180.95	614.05	0.1476	0.0069	0.0468
	NTN	177	190	205	10.5051	0.0131	181.64	**619.36**	0.0272	0.0095	**0.3511**
Host population	Potato	425	**442**	**611**	49.039	0.0613	181.04	616.95	0.2314	0.0210	0.0922
	Tobacco	100	255	302	45.678	0.0572	180.82	611.17	0.2176	0.0182	0.0836
	Other host	17	188	208	**61.765**	**0.0771**	**182.08**	**618.91**	**0.2889**	**0.0298**	**0.1031**
Geographical population	Iran	26	**250**	**297**	58.370	0.0728	**182.32**	618.67	0.2193	**0.0393**	**0.1792**
	Middle East	37	183	203	43.540	0.0551	164.95	588.05	0.0726	0.0549	0.756
	Far East	97	201	226	46.479	0.0584	168.23	608.77	0.0626	0.0620	0.990
	Africa	90	200	217	44.387	0.0556	168.73	608.27	0.0584	0.0602	**1.031**
	Europe	147	213	251	46.727	0.0584	169.49	610.51	0.0630	0.0620	0.984
	N. America	129	199	214	39.581	0.0494	169.31	613.69	0.0612	0.0498	0.813
	S. America	16	177	201	**65.275**	**0.0814**	182.11	**618.88**	**0.3162**	0.0301	0.095

S: number of segregation sites; η (Eta): total number of mutations; k: average number of nucleotide differences between sequences; π (pi): nucleotide diversity; SS: total number of synonymous sites analyzed; NS: total number of non-synonymous sites analyzed; π (s): synonymous nucleotide diversity; π (a): non-synonymous nucleotide diversity. Bold: indicate the highest value of each test.

Table 4. Genetic differentiation estimates for populations of potato virus Y

Population	Ka/Ks	Ks*	Kst*	Ks*,Kst* P-value	Z*	P-value	Snn	P-value	F_{ST}
O/C	0.097	3.18	0.056	0.000 ***	7.59	0.000***	1.000	0.000***	0.403
O/N	0.061	3.05	0.134	0.000 ***	7.47	0.000 ***	0.991	0.000 ***	0.722
O/NTN	0.098	2.72	0.224	0.000 ***	7.51	0.000 ***	0.993	0.000 ***	0.770
C/N	0.062	3.37	0.144	0.000 ***	5.10	0.000 ***	1.000	0.000 ***	0.581
C/NTN	0.088	2.48	0.198	0.000 ***	6.32	0.000 ***	1.000	0.000 ***	0.625
N/NTN	0.038	2.34	0.196	0.000 ***	6.45	0.000 ***	1.000	0.000 ***	0.592
Potato/tobacco	0.083	3.50	0.014	0.000 ***	10.20	0.000 ***	0.824	0.000 ***	0.051
Potato/other	0.096	3.58	0.005	0.002 **	9.77	0.002 **	0.935	0.001 **	0.075
Tobacco/other	0.088	3.42	0.021	0.000 ***	7.76	0.000 ***	0.899	0.000 ***	0.100

Ks, Kst*, Z* and Snn are test statistics of genetic differentiation [30]; F_{ST} examines the extent of genetic differentiation between geographical isolates; *: 0.01<P<0.05; **: 0.001<P<0.01; ***: P <0.001; ns: not significant.

of comparisons were supported by *P*-values less than 0.05 (Hudson 2000). (Tables 4 and 5).

First, our statistical analysis revealed significant genetic differences between all pairs of the PVY phylogenetic populations defined in this study (6 pairs; Table 4). The highest (0.722) and the lowest (0.403) F_{ST} value was found when comparing PVY^{O} population with PVY^{NTN} and PVY^{C} populations, respectively. This suggested the existence of a great genetic differentiation between all PVY phylogenetic populations.

Considering host populations, the null hypothesis of no genetic differentiation was rejected between all pairs (3 pairs) of the PVY host populations (populations of potato, tobacco and other hosts; Table 4). Non-potato PVY populations could also be specified by phylogenetic analysis as they were mostly placed in the PVY^{C} clade (Fig. 1). The F_{ST} values among PVY host populations ranged between 0.051 and 0.100 (Table 4). This finding concluded a moderate genetic differentiation among PVY host populations.

Table 5. Genetic differentiation estimates for geographical populations of potato virus Y

Population	Ks*	Kst*	Ks*,Kst* P-value	Z*	P-value	Snn	P-value	F_{ST}
Iran/Asia	3.619	0.003	0.033 *	9.148	0.0190 *	0.78007	0.9090 ns	0.011
Iran/Middle East	3.528	0.004	0.167 ns	6.477	0.0490 *	0.737	0.0000 ***	-0.005
Iran/Far East	3.486	0.01537	0.0030 **	7.833	0.0010 **	0.837	0.0000 ***	0.009
Iran/Africa	3.699	0.010	0.095 ns	5.46011	0.1000 ns	0.62838	0.1300 ns	0.064
Iran/Europe	3.517	0.007	0.044 *	7.540	0.0210 *	0.81567	0.0000 ***	0.018
Iran/N. America	3.136	0.044	0.000 ***	7.475	0.0000 ***	0.89423	0.0000 ***	0.127
Iran/S. America	3.874	0.004	0.164 ns	5.645	0.1790 ns	0.70417	0.0090 **	0.019
Middle East/Far East	3.391	0.016	0.002 **	8.033	0.001 **	0.792	0.000 ***	0.049
Middle East/Africa	3.333	0.006	0.045 *	7.949	0.007 **	0.803	0.000 ***	0.020
Middle East/Europe	3.433	0.006	0.020 *	8.707	0.003 **	0.867	0.000 ***	0.036
Middle East/N. America	3.212	0.021	0.001 **	8.458	0.001 **	0.903	0.000 ***	0.073
Middle East/S. America	3.467	0.012	0.085 ns	6.025	0.069 ns	0.751	0.035 *	0.046
Far East/Africa	3.381	0.015	0.000 ***	8.6825	0.000 ***	0.818	0.000 ***	0.016
Far East/Europe	3.439	0.007	0.005 **	9.257	0.000 ***	0.809	0.000 ***	0.005
Far East/N. America	3.295	0.040	0.000 ***	9.011	0.000 ***	0.853	0.000 ***	0.137
Far East/S. America	3.453	0.001	0.238 ns	7.674	0.156 ns	0.828	0.201 ns	-0.020
Africa/Europe	3.439	-0.00016	0.371 ns	7.341	0.2830 ns	0.85906	0.0060 **	-0.0004
Africa/N. America	3.012	0.044	0.000 ***	7.287	0.0000 ***	0.87686	0.0010 **	0.288
Africa/S. America	3.756	0.006	0.217 ns	4.992	0.2380 ns	0.60920	0.1520 ns	0.044
Europe/N. America	3.187	0.067	0.000 ***	8.246	0.0000 ***	0.88931	0.0000 ***	0.213
Europe/S. America	3.517	0.007	0.062 ns	7.384	0.055 ns	0.898	0.000 ***	0.019
N. America/S. America	3.103	0.033	0.000 ***	7.356	0.000 ***	0.880	0.000 ***	0.122

Ks, Kst*, Z* and Snn are test statistics of genetic differentiation [30]; F_{ST} examines the extent of genetic differentiation between geographical isolates; *: 0.01<P<0.05; **: 0.001<P<0.01; ***: P <0.001; ns: not significant.

Analysis of CP gene sequences using Kst*, Z and Snn test statistics revealed that most of the geographical PVY populations (including Asian, African, European, North American and South American PVY populations) were genetically distinct (Table 5). However, the highest F_{ST} values were obtained when North American PVY population was compared with African and European populations. The extent of genetic diversity between most of the geographical population pairs was little or moderate (F_{ST} <0.065). The exception was North American PVY population, which had great genetic differences with most of the other populations. For example the comparison of Iranian PVY population with the North American population and other geographical populations revealed F_{ST} values of 0.127 and at most 0.064, respectively (Table 5).

Based on these test statistics, geographical isolation may have played a role in PVY population structure especially in North America.

Considering F_{ST} values, the clear distinction of genetic populations of PVY can be achieved by phylogenetic analysis, host relevance and geographical distribution, respectively.

Examination of departure from neutrality

The patterns of nucleotide polymorphism in the CP gene of PVY isolates were estimated using Tajima's *D*, Fu and Li's *D** and *F** statistical tests (Table 6). The negative values of these test statistics would suggest for purifying selection and/or a recent population expansion (Tajima, 1989; Tsompana *et al.*, 2005). On the other hand, positive values would suggest for balancing selection and/or a reduction in population size.

Results showed that Tajima's *D*, Fu and Li's *D** and *F** statistics were negative for most of the phylogenetic and host populations of PVY. However, these negative values were only significant for PVY^{O} and PVY^{NTN} phylogenetic populations and PVY population isolated from potato (Table 6). Moreover, high level of haplotype diversity (Hd) and low rate of nucleotide diversity (π, except for PVY^{C}) were observed in the phylogenetic PVY populations. These findings suggested that these PVY populations may have been influenced by a recent population expansion.

Except the Asian PVY population, no significant departure from neutrality was found for PVY geographical populations according to Tajima's D, Fu and Li's D* and F* statistics (Table 6). Significantly negative test statistics obtained for the Asian population proposed the occurrence of a recent demographic expansion in this population. Also, high haplotype and high genetic diversities were observed in PVY geographical populations.

Recombination events

Searching for evidence of recombination events in the CP gene sequences showed that most of the PVY isolates were recombinants. In this regard 60% (15 out of 25) of the Iranian PVY isolates were found to be recombinants.

Our statistical analysis identified at least 13 recombination patterns in the CP gene sequences of PVY populations. (Table 7, Fig. 1).

Ten recombination patterns (events 1–10) appeared to be derived from parents from different PVY lineages (i.e. interlineage recombinants). In contrast, three recombination patterns (events 11–13) were found to be derived from parents within the same PVY lineage (i.e. intralineage recombinants) (Table 7).

Nine out of 13 recombination patterns were detected by at least four methods (events 1, 2, 3, 4, 5, 8, 9, 10, 12), hence

Table 6. Representation of parameter estimates and test statistics for demographic trends in potato virus Y populations

Population	Tajima's Da	Fu & Li's D*	Fu & Li's F*	Hd	π
O	-2.441**	-5.713**	-5.132**	0.997	0.030
C	-0.486ns	-0.443ns	-0.528ns	0.985	0.069
N	-0.069ns	-1.034ns	-0.856ns	1.000	0.033
NTN	-2.792***	-6.104**	-5.810	0.984	0.013
Potato	-1.555ns	-7.366**	-5.064**	0.994	0.061
Tobacco	-0.732ns	-2.278ns	-1.921ns	0.998	0.057
Other host	0.016ns	-0.474ns	-0.385ns	0.985	0.077
Iran	-1.072ns	-2.269ns	-2.220ns	1.000	0.072
Middle East	-0.392ns	-2.339ns	-1.965ns	0.995	0.055
Far East	0.197ns	-1.021ns	-0.593 ns	0.988	0.058
Asia	-1.582ns	-6.566**	-4.837**	0.996	0.063
Africa	0.588ns	0.252ns	0.393ns	0.974	0.058
Europe	-0.0523ns	-1.332ns	-0.959ns	0.991	0.058
N. America	-0.809ns	-1.711ns	-1.602ns	0.954	0.042
S. America	0.337	0.502	0.526	1.000	0.081

ns: non-significant; Hd: haplotype diversity; π: nucleotide diversity per site; **: 0.001<P<0.01.

Table 7. Characteristics of recombination events detected in isolates of potato virus Y

Event	Frequency	Recombination site	Recombinant length	Parental sublineage[a]	Recombination detected programs[b]	P-value[c]
1	131	352-801	449	O*N	**R.G.**B.**M.C.S.**P.L.**Se.**	1.17E-08
2	30	331-555	224	O*Nw	R.**G.**B.**M.C.S.**P.L.**Se.**	1.79E-07
3	153	147-342	195	Nw*NO	**R.G.**B.**M.C.S.**P.L.**Se.**	7.55E-03
4	4	16-164	148	C*O	R.G.B.**M.C.S.**P.L.**Se.**	5.51E-03
5	53	410-607	197	O*NTN	R.G.B.**M.C.S.**P.L.**Se.**	2.51E-02
6	2	496-636	140	C*O	R.G.B.M.C.S.P.L.**Se.**	1.09E-02
7	5	8-344	336	C*NTN	R.G.B.M.C.**S.**P.L.**Se.**	2.05E-02
8	141	345-745	400	O*No	**R.G.B.M.C.S.**P.L.**Se.**	1.33E-09
9	103	166-708	542	O*NTN	**R.G.**B.**M.C.S.**P.L.**Se.**	4.65E-07
10	249	351-751	400	O*NTN	**R.G.**B.**M.C.**S.P.L.**Se.**	6.90E-08
11	1	1-439	438	NTN*NTN	R.**G.**B.**M.C.**S.P.L.**Se.**	1.31E-03
12	68	606-791	185	NTN*NTN	**R.G.**B.**M.C.**S.P.L.**Se.**	3.54E-02
13	1	1-345	344	NTN*NTN	R.**G.**B.**M.C.**S.P.L.Se.	1.70E-02

(a): Parental sublineage estimated by RDP Program; (b): Abbreviations used for recombination detection programs; R.: RDP, G.: GENECONV, B.: Bootscan, M.: Maxchi, C.: Chimaera, S.: SiSscan, P.: PhylPro, L.: Lard, Se.: 3Seq. Recombination-detecting programs representing significant signal showed in bold; (c): The greatest P-value calculated by the program for the recombination event. The highest reported P-value for the program showed in underlined in RDP4.

considered as "significant recombination events". The four other recombination patterns were detected by less than four methods (events 6, 7, 11, 13), hence considered as "tentative recombination events" (Table 7).

Interestingly, our analysis revealed that most of the PVY recombinants were classified within the N:NTN clade with no preference between N-Europe and North American subclades. Contrary to this, many of the non-recombinant PVY strains were found within the $PVY^{O:C}$ clade (Fig. 1). PVY^{O} and PVY^{C} clades contained only one and two recombination patterns, respectively (Table 7). The recombination pattern in the PVY^{C} lineage seemed to be derived from parents from PVY^{O} and PVY^{C} clades. On the other hand, the parents of PVY recombinants in PVY^{O} were from the same lineage or from PVY^{O} and PVY^{C} lineages (Fig. 1).

Six out of 13 recombination patterns (including patterns of 1, 2, 5, 7, 8, 11) were found among the Iranian PVY population. The recombination pattern No. 7 was the most prevalent pattern within the Iranian PVY population (Fig. 1, Table 7).

Biological characterization of the PVY isolates

In order to compare the symptoms of PVY strains from divergent evolutionary lineages, four Iranian isolates representative of the PVY^{O} clade (isolates 21 and 221) and PVY^{NTN} clade (isolates 61 and IRAN) were inoculated to tobacco, pepper, tomato and *Ch. quinoa* plants.

Two weeks post-inoculation, all isolates induced vein clearing and mottling in *N. tabacum* cv. Xanthi and mottling and leaf rugosity in *L. esculentum*. In addition to mottling, *C. annuum*, developed the symptoms of vein necrosis by isolates 221 and IRAN, vein banding by isolate 21 and rogusity by isolate 61 (Table 8). None of the isolates infected *Ch. quinoa* Wild. Except *Ch. quinoa*, PVY was detected in all the inoculated plants using RT-PCR analysis.

Table 8. Biological characterization of Iranian potato virus Y isolates reported in this study

Isolate strain	21 PVY_{O}	221 PVY_{O}	IRAN PVY_{NTN}	61 PVY_{NTN}
Capsicum annuum	M, VB	M, VN	M, VN	R, M
Nicotiana tabacum cv. Xanthi	VC, M	VC, M	VC, M	VC, M
Lycopersicum esculentum	M, R	M, R	M, R	M, R

M: Mottle; R: Rugosity; VC: Vein clearing; VN: Vein necrosis; VB: Vein banding.

Discussion

PVY is a typical example of RNA viruses that benefits the high mutation and recombination rates for adaptation and survival in different hosts and various environments (Boonham *et al.*, 2002; Hu *et al.*, 2009a). Several studies have attempted to determine the genetic structure of PVY populations in Europe, North America and Japan (Glais *et al.*, 2002; Lorenzen *et al.*, 2006; Schubert *et al.*, 2007). In this work, we used the CP gene sequences to compare the genetic structure of PVY populations in Iran with the world

populations. We used CP gene, for which a large number of international sequence data exists in GenBank.

Phylogenetic analysis using CP gene sequence revealed the polyphyletic relationships of PVY strains reported from 34 countries across the world. Results showed that all the PVY isolates could be arranged within 2 main clades designated as $PVY^{O:C}$ and $PVY^{N:NTN}$. $PVY^{O:C}$ and $PVY^{N:NTN}$ could be further differentiated into 2 subclades designated as PVY^{O} and PVY^{C}, and N-Europe and North America, respectively (Glais *et al.*, 2002; Moury *et al.*, 2002; Fanigliulo *et al.*, 2005; Lorenzen *et al.*, 2006). The phylogenetic grouping of the PVY isolates was not fully consistent with their geographical distribution. For example, Iranian PVY strains were found to be distributed throughout the tree, suggesting multiple introductions of PVY virus isolates to Iran. Several PVY strains within $PVY^{O:C}$ and $PVY^{N:NTN}$ clades formed a star-like phylogeny, most probably as the consequence of recent emergences with minimal selection. Similar star phylogenies have been previously reported in the genetic structure of several virus populations including cucumber mosaic virus, pepino mosaic virus, wheat streak mosaic virus and turnip mosaic virus (Roossinck *et al.*, 1999; Pagan *et al.*, 2006; Dwyer *et al.*, 2007; Tomitaka *et al.*, 2007). Also some of the PVY sub-clades in the $PVY^{N:NTN}$ clades had low bootstrap value. This is most probably due to the recombination events as also considered by Ohshima *et al.* (2007).

Since the identification of first PVY recombinant strain from Hungary (PVY^{NTN}) in 1980s, several others have been reported in other countries (Le Romancer *et al.*, 1994; Glais *et al.*, 1996; Boonham *et al.*, 2002). In the past decade, many studies have shown the global prevalence of PVY recombinants (Chrzanowska, 1991; Glais *et al.*, 2002; Lorenzen *et al.*, 2006; Hu *et al.*, 2009b). In Europe, the non-recombinant strains of the PVY^{O} clade have been largely replaced by the recombinant strain of PVY^{NTN} (Boonham *et al.*, 2002). However, in North America, the PVY^{O} strain has remained the predominant strain in potato (Piche *et al.*, 2004; Baldauf *et al.,* 2006; Karasev *et al.*, 2008; Gray *et al.*, 2010). Our analysis revealed that PVY^{O} and PVY^{NTN} were predominant strains in Iran. Comparing results of this study with those published earlier (Pourrahim *et al.*, 2007; Hosseini *et al.*, 2011) suggested that during the past decade the prevalence of PVY^{NTN} was significantly increased within Iranian PVY population.

The genetic diversity of the CP gene was varied among different clades of the PVY phylogenetic tree, thereby some clades contained more daughter subclades than others. The combination of high haplotype diversity and low genetic diversity in phylogenetic PVY populations could be explained by a recent population expansion after a genetic bottleneck (Grant and Bowen, 1998; Tsompana *et al.*, 2005). However, relatively long terminal branches of the PVY^{C} isolates may suggest that they have accumulated mutations over a long time. All statistical analyses clearly suggested the significant genetic differences between pairs of the main PVY phylogenetic clades. The extent of this genetic variation between phylogenetic populations was considerable as shown by F_{ST} values. Also, our results revealed that PVY genetic diversity may have resulted from host adaptation, which is in agreement with previous reports (Schubert *et al.*, 2007; Ogawa *et al.*, 2008). However, the extent of the genetic variability between pairs of host PVY populations was low (F_{ST} <0.1).

Besides host adaptation, some of the PVY genetic diversity was found to have resulted from geographical distribution. In this regard, the majority of the geographical PVY population pairs could be genetically differentiated. However, the extent of genetic differences between most of the geographical populations was low. Contrasting to Cuevas *et al.* (2012), our analysis concluded the distinction of Middle East PVY population from populations of Europe and Far East. According to F_{ST} values, Iranian PVY isolates displayed the highest genetic differentiation from isolates of North America (F_{ST} = 0.127). This finding parallels the phylogenetic analysis, as no Iranian PVY isolates grouped within the North American clade. Our statistical analysis showed that North American PVY population had the highest genetic distance with other geographical populations (F_{ST} >0.07). Altogether, findings of this study confirmed that geographical distribution and host adaptation have played a role in shaping genetic diversity of the PVY populations.

The extent of the selection pressure acting on genes can be evaluated by computing the dN/dS, the ratio of the non-synonymous to the synonymous polymorphisms (Garcia-Arenal *et al.*, 2001). The global dN/dS ratio of the PVY CP gene was less than 1 (~0.08) revealing the exertion of high negative selection pressure on this gene. The negative selection pressure restricts the variability PVY CP gene. Different dN/dS values found on the CP genes of distinct PVY populations may suggest that these populations have experienced different constrains.

Analysis of dN/dS revealed that the CP gene sequence of PVY^{C} population experienced higher negative selection pressure compared to that of PVY^{NTN} and PVY^{O} strains. This is consistent with low CP gene sequence diversity (π = 0.069) of the PVY^{C} population as also considered by others (Schubert *et al.*, 2007; Ogawa *et al.*, 2008). The low diversity of CP gene in PVY^{C} could also be explained by Muller's ratchet phenomenon, i.e. the stochastic loss of genotypes (Garcia-Arenal *et al.*, 2001).

In most cases when phylogenetic, host or geographical PVY populations were considered, Tajima's D and Fu and Li's D and *F*-test yielded negative values. The negative values obtained in these tests in combination with high haplotype diversity and overall low nucleotide diversity of CP gene suggested that PVY phylogenetic population has undergone a recent population expansion or background selection.

High rates of haplotype and genetic diversity in the PVY host and geographical populations could reduce the risk of extinction and points to the evolutionary potential of these populations for adaptation into the diverse environments.

Recombination has been considered as an important source of genetic variation in potyviruses (Gibbs and Ohshima, 2010). For example, 76% of isolates of the potyvirus turnip mosaic virus were reported to be recombinant (Ohshima *et al.*, 2007). Our analysis also showed the occurrence of recombination events in the CP gene of PVY isolates as also considered by others (Glais *et al.*, 2002; Moury *et al.*, 2002; Lorenzen *et al.*, 2006). Besides the six recombination patterns identified in the earlier studies (Moury *et al.*, 2002; Fanigliulo *et al.*, 2005; Lorenzen *et al.*, 2006), we identified seven new patterns of recombination in the CP gene of the PVY strains. Our findings also conclude that recombination in the CP gene is a common and frequent force driving the PVY evolution.

Biological features of PVY strains have been described in several previous studies (Chrzanowska, 1991; Le Romancer *et al.*, 1994; dAquino *et al.*, 1995; Boonham *et al.*, 2002; Baldauf *et al.,* 2006). Tomato and tobacco plants can be infected by most PVY isolates (Stobbs *et al.*, 1994). Commonly, PVY isolates of pepper cannot infect potato plants systemically (McDonald and Kristjansson, 1993; dAquino *et al.*, 1995). Whereas members of the PVY^{O} and PVY^{C} populations could induce mottling in pepper plants (McDonald and Kristjansson, 1993; Valkonen *et al.*, 1996), PVY^{NNP} reported from Italy has been the only isolate inducing vein necrosis in pepper (Fanigliulo *et al.*, 2005). Our experimental data revealed that the two Iranian PVY isolates (PVY-Iran and PVY-221) could also induce vein necrosis symptom in pepper, hence they were biologically similar to the PVY^{NNP} isolate (dAquino *et al.*, 1995).

In this study we described the structure and dynamics of the PVY populations in the world, with a focus on the PVY isolates from Iran. Evolutionary studies showed that PVY populations could be represented into discrete lineages. Geographical origin, mutation, recombination and host adaptation were the main sources of genetic variability to shape the population structure of PVY. Strong negative selection pressure on the CP gene favors the preservation of wild type strains in the nature. These findings argue that the CP gene is an effective indicator to study the genetic diversity and evolution of PVY populations. Iranian PVY population displayed high haplotype and nucleotide diversities in the CP gene. The high nucleotide diversity within Iranian PVY population is probably due to the introduction of distant strains belonging to PVY^{O}, PVY^{C} and N-Europe phylogenetic populations, as also shown in the phylogenetic tree. Altogether, phylogenetic, biological and sequence diversity analyses of the CP gene suggested that Iranian PVY isolates constitute a diverse population. Our analysis also proposes that the PVY^{O} strains have been largely replaced by the recombinant strains of PVY^{NTN} within the Iranian PVY population during the last decade. The genetic variability of the PVY populations reported in this study may provide foundation to improve control strategies of PVY in different crops.

References

Baldauf PM, Gray SM, Perry KL (2006): Biological and serological properties of Potato virus Y isolates in northeastern United States potato. Plant Dis. 90, 559–566. https://doi.org/10.1094/PD-90-0559

Bermingham E, Moritz C (1998): Comparative phylogeography: concepts and application. Mol. Ecol. 7, 367–369. https://doi.org/10.1046/j.1365-294x.1998.00424.x

Boonham N, Walsh K, Hims M, Preston S, North J, Barker I (2002): Biological and sequence comparisons of Potato virus Y isolates associated with potato tuber necrotic ringspot disease. Plant Pathol. 51, 117–126. https://doi.org/10.1046/j.1365-3059.2002.00681.x

Chikh Ali M, Maoka T, Natsuaki T, Natsuaki KT (2010): PVYNTN-NW, a novel recombinant strain of Potato virus Y predominating in potato fields in Syria. Plant Pathol. 59, 31–41. https://doi.org/10.1111/j.1365-3059.2009.02174.x

Chrzanowska M (1991): New isolates of the necrotic strain of potato virus Y (PVYN) found recently in Poland. Potato Res. 34, 179–182. https://doi.org/10.1007/BF02358039

Chung BY, Miller WA, Atkins JF, Firth AE (2008): An overlapping essential gene in the Potyviridae. Proc. Natl. Acad. Sci. USA 105, 5897–5902. https://doi.org/10.1073/pnas.0800468105

Clark CA, Abad JA, Cuellar WJ (2012): Sweet potato viruses: 15 years of progress on understanding and managing complex diseases. Plant Dis. 96, 168–185. https://doi.org/10.1094/PDIS-07-11-0550

Clark MF, Adams AN (1977): Characteristics of the microplate method of enzyme-linked immunosorbent assay for the detection of plant viruses. J. Gen. Virol. 34, 475–483. https://doi.org/10.1099/0022-1317-34-3-475

Cuevas JM, Delaunay A, Visser JC, Bellstedt DU, Jacquot E, Elena SF (2012): Phylogeography and molecular evolution of potato virus Y. PLoS ONE 7, e37853. https://doi.org/10.1371/journal.pone.0037853

dAquino L, Dalmay T, Burgyan J, Ragozzino A, Scala F (1995): Host range and sequence analysis of an isolate of potato virus Y inducing veinal necrosis in pepper. Plant Dis. 79, 1046–1050. https://doi.org/10.1094/PD-79-1046

Dwyer GI, Gibbs MJ, Gibbs AJ, Jones RAC (2007): Wheat streak mosaic virus in Australia: relationship to isolates from the Pacific Northwest of the USA and its dispersion via seed transmission. Plant Dis. 91, 164–170. https://doi.org/10.1094/PDIS-91-2-0164

Elena SF, Bedhomme S, Carrasco P, Cuevas JM, de laIglesia F, Lafforgue G, Lalic J, Prosper A, Tromas N, Zwart MP (2011): The evolutionary genetics of emerging plant RNA viruses.

Mol. Plant Microbe Interact. 24, 287–293. https://doi.org/10.1094/MPMI-09-10-0214

Fanigliulo A, Comes S, Pacella R, Harrach B, Martin DP, Crescenzi A (2005): Characterization of Potato virus Y nnp strain inducing veinal necrosis in pepper: a naturally occurring recombinant strain of PVY. Arch. Virol. 150, 709–720. https://doi.org/10.1007/s00705-004-0449-x

Fauquet CM, Mayo MA, Maniloff J, Desselberger U, Ball LA (2012): Virus Taxonomy. Eighth Report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, San Diego, California.

Fu YX, Li WH (1993): Statistical tests of neutrality of mutations. Genetics 133, 693–709.

Garcia-Arenal F, Fraile A, Malpica JM (2001): Variability and genetic structure of plant virus populations. Annu. Rev. Phytopathol. 39, 157–186. https://doi.org/10.1146/annurev.phyto.39.1.157

Gibbs AJ, Ohshima K (2010): Potyviruses and the digital revolution. Annu. Rev. Phytopathol. 48, 205–223. https://doi.org/10.1146/annurev-phyto-073009-114404

Glais L, Tribodet M, Kerlan C (2002): Genomic variability in Potato potyvirus Y (PVY): evidence that PVYN-W and PVYNTN variants are single to multiple recombinants between PVYO and PVYN isolates. Arch. Virol. 147, 363–378. https://doi.org/10.1007/s705-002-8325-0

Glais L, Kerlan C, Tribodet M, Tordo V, Robaglia C, Astier-Manifacier S (1996): Molecular characterization of potato virus YN isolates by PCR-RFLP. Eur. J. Plant Pathol. 102, 655–662. https://doi.org/10.1007/BF01877246

Grant WS, Bowen BW (1998): Shallow population histories in deep evolutionary lineages of marine fishes: insights from sardines and anchovies and lessons for conservation. J. Heredity 89, 415–426. https://doi.org/10.1093/jhered/89.5.415

Gray SM, DeBoer SH, Lorenzen J, Karasev AV, Whitworth J, Nolte P, Singh RP, Boucher A, Xu H (2010): Potato virus Y: A significant and evolving threat to potato crops in the United States and Canada. Plant Dis. 94, 1384–1397. https://doi.org/10.1094/PDIS-02-10-0124

Heath L, Vander Walt E, Varsani A, Martin DP (2006): Recombination patterns in aphthoviruses mirror those found in other picornaviruses. J. Virol. 80, 11827–11832. https://doi.org/10.1128/JVI.01100-06

Hillung J, Elena S, Cuevas J (2013): Intra-specific variability and biological relevance of P3 N-PIPO protein length in potyviruses. BMC Evol. Biol. 13, 249. https://doi.org/10.1186/1471-2148-13-249

Hosseini A, Massumi H, Heydarnejad J, Varsani A, Hosseini Pour A (2011): characterization of potato virus Y isolates from Iran. Virus Genes 42, 128–140. https://doi.org/10.1007/s11262-010-0546-8

Hu X, Karasev AV, Brown CJ, Lorenzen JH (2009a): Sequence characteristics of Potato virus Y recombinants. J. Gen. Virol. 90, 3033–3041. https://doi.org/10.1099/vir.0.014142-0

Hu X, Meacham T, Ewing L, Gray SM, Karasev AV (2009b): A novel recombinant strain of Potato virus Y suggests a new viral genetic determinant of vein necrosis in tobacco. Virus Res. 143, 68–76. https://doi.org/10.1016/j.virusres.2009.03.008

Hudson RR (2000): A new statistic for detecting genetic differentiation. Genetics 155, 2011–2014.

Jones RA (2009): Plant virus emergence and evolution: origins, new encounter scenarios, factors driving emergence, effects of changing world conditions, and prospects for control. Virus Res. 141, 113–130. https://doi.org/10.1016/j.virusres.2008.07.028

Karasev AV, Gray SM (2013): Continuous and emerging challenges of Potato virus Y in potato. Annu. Rev. Phytopathol. 51, 571–586. https://doi.org/10.1146/annurev-phyto-082712-102332

Karasev AV, Meacham T, Hu X, Whitworth J, Gray SM, Olsen N, Nolte P (2008): Identification of PVY strains associated with tuber damage during an outbreak of PVY in potato in Idaho. Plant Dis. 92, 1371–1371. https://doi.org/10.1094/PDIS-92-9-1371A

Le Romancer M, Kerlan C, Nedellec M (1994): Biological characterization of various geographical isolates of Potato virus Y inducing superficial necrosis on potato tubers. Plant Pathol. 43, 138–144. https://doi.org/10.1111/j.1365-3059.1994.tb00563.x

Libardo P, Rozas J (2009): DnaSP V5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451–1452. https://doi.org/10.1093/bioinformatics/btp187

Lorenzen JH, Meacham T, Berger PH, Shiel PJ, Crosslin JM, Hamm PB, 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. https://doi.org/10.1007/s00705-005-0707-6

Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P (2010): RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics 26, 2462–2463. https://doi.org/10.1093/bioinformatics/btq467

McDonald JG, Kristjansson GT (1993): Properties of strains of potato virus YN in North America. Plant Dis. 77, 87–89. https://doi.org/10.1094/PD-77-0087

Moury B, 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. https://doi.org/10.1093/molbev/msr105

Moury B, Morel C, Johansen E, 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. https://doi.org/10.1099/0022-1317-83-10-2563

Nie X, Singh RP (2003): Evolution of North American PVY(NTN) strain Tu 660 from local PVY(N) by mutation rather than recombination. Virus Genes 26, 39–47. https://doi.org/10.1023/A:1022326021195

Ogawa T, Tomitaka Y, Nakagawa A, 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. https://doi.org/10.1016/j.virusres.2007.09.010

Ohshima K, Tomitaka Y, Wood JT, Minematsu Y, Kajiyama H, Tomimura K, Gibbs AJ (2007): Patterns of recombination in turnip mosaic virus genomic sequences indicate hotspots of recombination. J. Gen. Virol. 88, 298–315. https://doi.org/10.1099/vir.0.82335-0

Pagan I, Cordoba-Selles MC, Martınez-Priego L, Fraile A, Malpica JM, Jorda C, Garcia-Arenal F (2006): Genetic structure of the population of Pepino mosaic virus infecting tomato crops in Spain. Phytopathol. 96, 274–279. https://doi.org/10.1094/PHYTO-96-0274

Piche LM, Singh RP, Nie X, Gudmestad NC (2004): Diversity among Potato virus Y isolates obtained from potatoes grown in the United States. Phytopathol. 94, 1368–1375. https://doi.org/10.1094/PHYTO.2004.94.12.1368

Pourrahim R, Farzadfar S, Golnaraghi AR, Ahoonmanesh A (2007): Incidence and distribution of important viral pathogens in some Iranian potato fields. Plant Dis. 91, 609–615. https://doi.org/10.1094/PDIS-91-5-0609

Roossinck MJ, Zhang L, Hellwald KH (1999): Rearrangements in the 5` nontranslated region and phylogenetic analyses of Cucumber mosaic virus RNA 3 indicate radial evolution of three subgroups. J. Virol. 73, 6752–6758.

Schubert J, Fomitcheva V, Sztangret-Wisniewska J (2007): Differentiation of Potato virus Y strains using improved sets of diagnostic PCR-primers. J. Virol. Methods 140, 66–74. https://doi.org/10.1016/j.jviromet.2006.10.017

Shukla DD, Ward CW, Brunt AA (1994): The Potyviridae. CAB International, Wallingford, UK.

Stobbs LW, Poysa V, Van Schagen JG (1994): Susceptibility of cultivars of tomato and pepper to a necrotic strain of potato virus Y. Can. J. Plant Pathol. 16, 43–48.

Tajima F (1983): Evolutionary relationship of DNA sequences in finite populations. Genetics 105, 437–460.

Tajima F (1989): Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 585–595.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011): MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739. https://doi.org/10.1093/molbev/msr121

Tomitaka Y, Yamashita T, Ohshima K (2007): The genetic structure of populations of Turnip mosaic virus in Kyushu and central Honshu. J. Gen. Plant Pathol. 73, 197–208. https://doi.org/10.1007/s10327-007-0003-2

Tsompana M, Abad J, Purugganan M, Moyer JW (2005): The molecular population genetics of the Tomato spotted wilt virus (TSWV) genome. Mol. Ecol. 14, 53–66. https://doi.org/10.1111/j.1365-294X.2004.02392.x

Valkonen JPT, Kyle MM, Slack SA (1996): Comparison of resistance to potyviruses within Solanaceae: infection of potatoes with tobacco etch potyvirus and peppers with potato A and Y potyviruses. Ann. App. Biol. 129, 25–38. https://doi.org/10.1111/j.1744-7348.1996.tb05728.x

Visser JC, Bellstedt DU (2009): An assessment of molecular variability and recombination patterns in South African isolates of potato virus Y. Arch. Virol. 154, 1891–1900. https://doi.org/10.1007/s00705-009-0525-3