

# *Complete genome sequence of a novel Plum pox virus strain W isolate determined by 454 pyrosequencing*

**Anna Sheveleva, Anna Kudryavtseva,  
Anna Speranskaya, Maxim Belenikin,  
Natalia Melnikova & Sergei Chirkov**

## **Virus Genes**

ISSN 0920-8569

Volume 47

Number 2

Virus Genes (2013) 47:385–388

DOI 10.1007/s11262-013-0946-7



**Your article is protected by copyright and all rights are held exclusively by Springer Science +Business Media New York. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at [link.springer.com](http://link.springer.com)".**

# Complete genome sequence of a novel *Plum pox virus* strain W isolate determined by 454 pyrosequencing

Anna Sheveleva · Anna Kudryavtseva ·  
 Anna Speranskaya · Maxim Belenikin ·  
 Natalia Melnikova · Sergei Chirkov

Received: 11 February 2013 / Accepted: 19 June 2013 / Published online: 28 June 2013  
 © Springer Science+Business Media New York 2013

**Abstract** The near-complete (99.7 %) genome sequence of a novel Russian *Plum pox virus* (PPV) isolate Pk, belonging to the strain Winona (W), has been determined by 454 pyrosequencing with the exception of the thirty-one 5'-terminal nucleotides. This region was amplified using 5'RACE kit and sequenced by the Sanger method. Genomic RNA released from immunocaptured PPV particles was employed for generation of cDNA library using TransPlex Whole transcriptome amplification kit (WTA2, Sigma-Aldrich). The entire Pk genome has identity level of 92.8–94.5 % when compared to the complete nucleotide sequences of other PPV-W isolates (W3174, LV-141pl, LV-145bt, and UKR 44189), confirming a high degree of variability within the PPV-W strain. The isolates Pk and LV-141pl are most closely related. The Pk has been found in a wild plum (*Prunus domestica*) in a new region of Russia indicating widespread dissemination of the PPV-W strain in the European part of the former USSR.

**Electronic supplementary material** The online version of this article (doi:10.1007/s11262-013-0946-7) contains supplementary material, which is available to authorized users.

A. Sheveleva · S. Chirkov (✉)  
 Department of Virology, Biological Faculty, Lomonosov  
 Moscow State University, Leninskie Gory 1/12,  
 Moscow 119991, Russia  
 e-mail: s-chirkov1@yandex.ru

A. Kudryavtseva · A. Speranskaya · M. Belenikin ·  
 N. Melnikova  
 Engelhardt Institute of Molecular Biology, Russian Academy of  
 Sciences, Vavilov str. 32, Moscow 119991, Russia

A. Speranskaya  
 Department of Higher Plants, Biological Faculty, Lomonosov  
 Moscow State University, Leninskie Gory 1/12,  
 Moscow 119991, Russia

**Keywords** *Plum pox virus* · Strain Winona · Complete genome sequence · 454 pyrosequencing

*Plum pox virus* (PPV; genus *Potyvirus*; family *Potyviridae*) is the most important viral pathogen of stone fruit crops [1]. The PPV genome is typical of potyviruses and consists of a single-stranded positive sense RNA approximately 9.8 kb in length with a 5'-terminal viral genome-linked protein (VPg) and a 3' poly (A) tail. The genome is bicistronic with two open reading frames (ORFs). A long ORF, flanked with the 5'- and 3'-non-coding regions (NCR), is translated into approximately 355 kDa polyprotein processed by virus-encoded proteases to ten functional proteins, namely P1, HC-Pro, P3, 6K1, CI, 6K2, VPg, NIa-Pro, NIb, and coat protein (CP) [2, 3]. A short overlapping ORF, called PIPO, is encoded within the P3 cistron and expressed by −1 ribosomal frameshifting as a product fused with the N-terminal region of P3 [4]. Based on sequence differences of genomic RNAs, phylogenetic analysis, and antigenic properties, seven PPV strains (PPV-D, PPV-M, PPV-Rec, PPV-C, PPV-EA, PPV-W, and PPV-T) are recognized to date [2, 5]. The strains D, M, and Rec are widespread in Europe while EA, C, W, and T are likely to be minor groups in restricted regions [5–7].

*Plum pox virus*-W is a rare, little-studied, and apparently the most variable PPV strain of the seven recognized strains of the virus [8–14]. The complete genome sequences of four PPV-W isolates are currently available: W3174 (GenBank accession number AY912055, [9]), LV-141pl (HQ670746, [13]), LV-145bt (HQ670748, [13]), and UKR44189 (JN596110, [10]). Also, the genomes of some other PPV-W isolates from Latvia and Russia were partially sequenced [11–13], and a number of unpublished sequences deposited in GenBank]. Molecular characterization of new PPV-W

isolates is important for understanding of its evolution and the degree of genetic variability.

In this study we determined the complete genome sequence of a novel Russian isolate Pk, that belongs to PPV-W group, by 454 pyrosequencing, and we compared the genome sequence of Pk with the sequences of other PPV-W isolates available in the GenBank database.

The isolate Pk was found in a symptomatic wild *Prunus domestica* tree growing in the Tver region of Russia. The virus was detected in the leaf extract using the polyclonal double antibody sandwich ELISA with a reagent set SRA 31505 (Agdia, USA) and using immunocapture-reverse transcription-polymerase chain reaction (IC-RT-PCR) with universal 3'-NCR-specific primers [15] as described previously [12]. The Pk was identified as a member of the PPV-W strain using RT-PCR with the PPV-W-specific primers 3174-SP-F3/3174-SP-R1 [16] and W8328F/W8711R [13]. The isolate 1410 [11] served as a PPV-W positive control (data not shown). The result of the strain identification by RT-PCR was confirmed using CP gene sequencing (JQ970440).

Genomic RNA released from immunocaptured PPV particles was employed for the synthesis and amplification of cDNA library using TransPlex Whole transcriptome amplification kit (WTA2, Sigma-Aldrich, USA) according to the manufacturer's instructions. The 454 pyrosequencing and analysis of generated sequence data were performed using GS Junior System (Roche). The 37572 raw reads (average length of 292 nucleotides) were mapped to the PPV isolate LV-145bt (HQ670748) as a reference sequence using Newbler Software (version 2.7). The 18968 reads (50.4 %) generated one contig with a length of 9758 nucleotides (nt) (mean 500-fold coverage). Thus, the assembled contig covered 99.7 % of the scaffold genome with the exception of the first thirty-one 5'-terminal nt. The remaining 5'-region was amplified using 5'RACE kit (Invitrogen, USA), following the manufacturer's instructions, and sequenced by Sanger sequencing method in Evrogen facilities (Moscow, Russia). The genome sequence of the Pk has been deposited to GenBank under accession number KC347608.

The complete Pk genome consists of 9789 nt, excluding the 3'-terminal poly (A) tail. Typical of PPV, a long ORF starts at an AUG (nt positions 147–149) and terminates at an UAG stop codon (nt positions 9570–9572). Thus, this ORF consists of 9,423 nt and potentially encodes a polyprotein of 3,141 amino acids (aa) with a calculated molecular weight (MW) of 355.6 kDa. The ORF is flanked with the 5'-NCR and 3'-NCR composed of 146 and 217 nt, respectively. The PIPO-1 ORF starts at a conservative G<sub>2</sub>A<sub>6</sub> motif in the P3 cistron (nt position 2906) and terminates at the nearest UGA stop codon in PIPO in-frame (nt positions 3224–3226). Thus, this gene consists of 106

codons. A predicted MW of the putative (N-ter) P3-PIPO fusion product is about 29 kDa.

The comparison of the complete genome sequences of Pk and other fully sequenced PPV-W isolates including their individual genes is presented in Table 1. The percentage of the nt sequences identity was calculated using Martinez-NW method [17] with MegAlign v.7.1.0 software. The entire Pk genome has nt identities of 92.8–94.5 % when compared with the complete nt sequences of the other isolates, confirming a high degree of genetic variability in the PPV-W group [12–14]. Apparently, the isolates Pk and LV-141pl are most closely related.

The 3'-NCR, PIPO, and C-terminal part of the CP are obviously the most conservative regions within the five PPV-W genomes. On the contrary, striking differences between Pk and other PPV-W isolates are revealed in the 5'-terminal sequences of the CP gene (the first 303 nt), encoding the N-terminal domain of the CP, and in the 6K2 and P1 genes. The N-terminal region of CP and P1 are known to be the most variable proteins among the potyviruses [3]. The lowest level of nt identity has been found among the P1, Hc-Pro, and 6K2 genes of the isolates Pk and W3174. This is probably due to the mosaic structure of the W3174 genome which contains PPV-M and PPV-D—specific recombinant fragments in P1/Hc-Pro and 6K2/VPg regions, respectively [13, 18]. No evidences of recombination events were observed when the complete genome sequence of Pk was analyzed using RDP program [19] (data not shown). The CP gene of the Pk is 993 nt long and the translated CP consists of 331 aa. It should be noted that

**Table 1** Percentage identities of the complete genomes and individual genes of *Plum pox virus* isolate Pk compared to other PPV-W isolates

Genome region	PPV-W isolates			
	W3174	LV-141pl	LV-145bt	UKR 44189
Complete genome	92.8	94.5	93.6	93.5
5'-NCR	95.2	93.2	92.5	91.8
P1	85.2	93.2	91.0	90.9
Hc-Pro	89.4	94.0	93.5	93.4
P3	95.0	95.5	93.0	93.2
PIPO	97.8	96.9	97.8	98.1
6K1	92.9	93.6	92.3	92.3
CI	94.9	94.0	93.9	94.0
6K2	84.3	95.0	91.2	91.8
VPg	93.6	94.1	94.5	94.5
NIa-Pro	95.2	96.0	94.7	94.4
NIb	94.2	95.2	94.0	93.9
CP (Nter/rest)	89.0/96.4	89.0/96.5	88.0/96.5	88.0/96.2
3'-NCR	96.8	97.7	97.7	97.3



the CP gene sequences determined by Sanger sequencing (JQ970440) and 454 pyrosequencing (KC347608, this work) were identical. Phylogenetic analysis (Fig. 1) confirmed the homology analysis results (Table 1) and allowed to assign the isolates Pk/LV-141pl, UKR44189/LV-145bt, and W3174 to the independent evolutionary lineages. These results were similar to the previous results [10, 13].

The PPV polyprotein is processed into mature products by virus-encoded proteases P1, Hc-Pro, and NIa-Pro [2, 3, 22]. Analysis of the deduced Pk polyprotein sequence shows that the NIa-Pro cleavage sites in CI/6K2 and N1b/CP junctions consist of the heptapeptides ECVHHQ/N and NIVVHQ/A, respectively, differing from their counterparts in other PPV-W isolates. The asparagine (N) at position +1 has not been described previously for PPV cleavage site between CI and 6K2 [3].

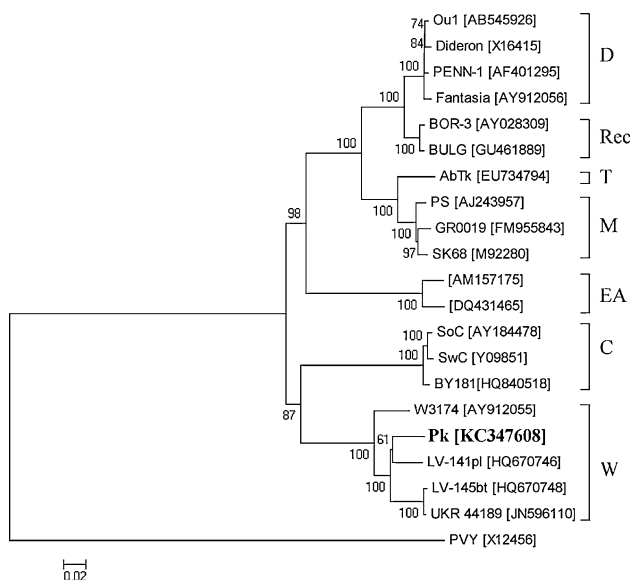
Variability of the PPV-W is two to eight times higher than those of other PPV strains [13]. Genetic diversity of the RNA viruses can be attributed in part to the low fidelity of viral RNA-dependent RNA polymerase (RdRp) [23]. To investigate the possible role of the PPV-W replicase (encoded by the N1b gene) in generation of the intrastrain molecular diversity, we have compared the deduced aa sequences of the replicases of five PPV-W isolates with those of several randomly selected members of other PPV

strains. The pairwise percentages of nt identity and aa similarity in the PPV-W N1b are 93.9–95.2 and 98.1–98.6 %, respectively, that is consistent with the levels of N1b diversity within strains D, M, and EA [24]. Eight conserved motifs have been identified in the RdRps of positive-strand RNA viruses; four of them, A–D (equivalent to the conserved motifs IV–VII of RdRp), form the catalytic center of the molecule [23–25]. The multiple alignment of the aa sequences of the PPV replicases shows that these motifs are very similar among the compared PPV isolates (Fig. S1). At the same time, multiple aa positions have been found to be specific for the PPV-W replicase only. Four of them (102, 110, 149, and 164) mapped to the N-terminal domain are involved in the binding of other proteins of the replicative complex. Valine/isoleucine substitution at position 358 is located within the magnesium coordination site of the catalytic domain. Seven aa positions (469, 483, 487, 490, 505, 508, and 509) have been found in the C-terminal region whose function is unknown [23].

The next-generation sequencing is a powerful tool for reliable detection and fast characterization of known and novel plant viruses [26–32]. Total, double-stranded, or small-interfering RNAs from infected plants and nucleic acids obtained from purified viral particles are commonly used as a starting material for these applications. In this work, genomic RNA released from immunocaptured particles has been successfully used for the synthesis and amplification of cDNA library, yielding up to 50 % virus-specific cDNA. This approach facilitates the procedures of pyrosequencing and contig assembly and it is also useful for rapid characterization of the genome of a known virus. The near-complete Pk genome (99.7 %) was sequenced with high confidence, owing to the mean 500× depth of sequence coverage. The reasons why the sequence of the first 31 nt was not determined by pyrosequencing are to be elucidated yet. To our knowledge, this is the first work in which the complete genome of PPV was defined using a next-generation sequencing approach.

It is also noteworthy that the Pk was discovered in yet another (Tver) region of Russia (approximately 250 km to the northwest from Moscow). To date, genetically diverse PPV-W isolates were found in Russia and Latvia and also in Canada and the USA in plums originated from Ukraine [8, 10–13]. No PPV-W isolates were found elsewhere. Taken together, this finding and the previous results [8–13] indicate widespread distribution of the PPV-W in the European part of the former USSR.

**Acknowledgments** This work was supported by the Russian Ministry of Education and Science (contract No. 16.552.11.7069). We thank Dr. Peter Ivanov (Department of Virology, Lomonosov Moscow State University) for critical reading of the manuscript.



**Fig. 1** Phylogenetic analysis of the complete nucleotide sequences of the Pk and 19 representative isolates belonging to the conventional PPV strains. The tree was built by the neighbor-joining method implemented to MEGA5 software [20]. The isolates were entitled by their names and accession numbers (Pk is in **bold**). Bootstrap values >60 % (1,000 bootstrap re-samplings) are indicated as *percentages* on the *branches*. The *scale bar* indicates the number of substitutions per residue. The complete nucleotide sequence of *Potato virus Y* [21] was used as the phylogenetic outgroup

## References

1. M. Cambra, D. Boscia, A. Myrta, G. Llacer, EPPO Bull. **36**, 202–204 (2006)
2. J.A. Garcia, M. Cambra, Plant Viruses **1**, 69–79 (2007)
3. B. Salvador, J.A. Garcia, C. Simon-Mateo, EPPO Bull. **36**, 229–238 (2006)
4. B.Y.-W. Chung, W.A. Miller, J.F. Atkins, A.E. Firth, Proc. Natl. Acad. Sci. USA **105**, 5897–5902 (2008)
5. C.U. Serce, T. Candresse, L. Svanella-Dumas, L. Krizbai, M. Gazela, K. Caglayan, Virus Res. **142**, 121–126 (2009)
6. T. Candresse, M. Cambra, EPPO Bull. **36**, 239–246 (2006)
7. D. James, M. Glasa, EPPO Bull. **36**, 247–250 (2006)
8. D. James, A. Varga, D. Thompson, S. Hayes, Plant Dis. **87**, 1119–1124 (2003)
9. D. James, A. Varga, Virus Res. **110**, 143–150 (2005)
10. V. Mavrodieva, D. James, K. Williams, S. Negi, A. Varga, R. Mock, L. Levy, Plant Dis. **97**, 44–52 (2012)
11. A. Sheveleva, E. Nemova, S. Chirkov, Acta Hortic. **899**, 49–55 (2011)
12. A. Sheveleva, P. Ivanov, Y. Prihodko, D. James, S. Chirkov, Plant Dis. **96**, 1135–1142 (2012)
13. M. Glasa, T. Malinowski, L. Predajna, N. Pupola, D. Deken, L. Michalczyk, T. Candresse, Phytopathology **101**, 980–985 (2011)
14. A. Sheveleva, P. Ivanov, Y. Prihodko, A. Varga, D. James, S. Chirkov, in *Book of Abstracts of the 22nd International Conference on Virus and other Transmissible Diseases of Fruit Crops*, ICVF, Rome, 2012, p. 35
15. L. Levy, A.A. Hadidi, EPPO Bull. **24**, 595–604 (1994)
16. D. James, A. Varga, Acta Hortic. **657**, 177–182 (2004)
17. H.M. Martinez, Nucl. Acids Res. **11**, 4629–4634 (1983)
18. A. Myrta, A. Varga, D. James, Arch. Virol. **151**, 1189–1198 (2006)
19. D.P. Martin, P. Lemey, M. Lott, V. Moulton, D. Posada, P. Le-feuvre, Bioinformatics **26**, 2462–2463 (2010)
20. K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, S. Kumar, Mol. Biol. Evol. **28**, 2731–2739 (2011)
21. C. Robaglia, M. Durand-Tardif, M. Tronchet, G. Boudazin, S. Astier-Manifacier, F. Casse-Delbart, J. Gen. Virol. **70**, 935–947 (1989)
22. J.A. Garcia, M.T. Martin, M.T. Cervera, J.L. Reichman, Virology **188**, 697–703 (1992)
23. E.K. O'Reilly, C. Cheng Kao, Virology **252**, 287–303 (1998)
24. M.-E. Aleman-Verdauger, C. Goudou-Urbino, J. Dubern, R.N. Beachy, C. Fauquet, J. Gen. Virol. **78**, 1253–1264 (1997)
25. E.V. Koonin, J. Gen. Virol. **72**, 2197–2206 (1991)
26. I.P. Adams, R.H. Glover, W.A. Monger, R. Mumford, E. Jackeviciene, M. Navalinskiene, M. Samuitiene, N. Boonham, Mol. Plant Pathol. **10**, 537–545 (2009)
27. M. Al Rwahnih, S. Daubert, D. Golino, A. Rowhani, Virology **387**, 395–401 (2009)
28. J.F. Kreuze, A. Perez, M. Untiveros, D. Quispe, S. Fuentes, I. Barker, R. Simon, Virology **388**, 1–7 (2009)
29. B. Coetzee, M.-J. Freeborough, H.J. Maree, J.-M. Celton, D.J.G. Rees, J.T. Burger, Virology **400**, 157–163 (2010)
30. M. Al Rwahnih, S. Daubert, J.R. Urbez-Torres, F. Cordero, A. Rowhani, Arch. Virol. **156**, 397–403 (2011)
31. P.E. Rodriguez Pardina, N. Bejerman, A.V. Luque, L. Di Feo, Virus Genes **45**, 593–595 (2012)
32. T. Candresse, A. Marais, C. Faure, P. Gentit, Phytopathology **103**, 293–298 (2013)