

Complete genome sequences of three Prunus-infecting nepoviruses: apricot latent ringspot virus, myrobalan latent ringspot virus and a novel virus from smooth stone peach (*Prunus mira* Koehne)

Maryam Khalili

INRAE Nouvelle-Aquitaine Bordeaux: Institut National de Recherche pour l'Agriculture l'Alimentation et l'Environnement Centre Nouvelle-Aquitaine Bordeaux

Thierry Candresse

INRAE Nouvelle-Aquitaine Bordeaux: Institut National de Recherche pour l'Agriculture l'Alimentation et l'Environnement Centre Nouvelle-Aquitaine Bordeaux

Chantal Faure

INRAE Nouvelle-Aquitaine Bordeaux: Institut National de Recherche pour l'Agriculture l'Alimentation et l'Environnement Centre Nouvelle-Aquitaine Bordeaux

Yoann Brans

CTIFL Centre de Lanxade

Armelle Marais

`armelle.marais-colombel@inrae.fr`

INRAE <https://orcid.org/0000-0003-2482-1543>

Research Article

Keywords: Prunus, high-throughput sequencing, Nepovirus, Secoviridae, ALRSV, MLRSV, stone fruit virus, *Prunus mira*

Posted Date: December 21st, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-3760896/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Archives of Virology on July 17th, 2024. See the published version at <https://doi.org/10.1007/s00705-024-06091-7>.

Complete genome sequences of three *Prunus*-infecting nepoviruses: apricot latent ringspot virus, myrobalan latent ringspot virus and a novel virus from smooth stone peach (*Prunus mira* Koehne)

M. Khalili^{1#}, T. Candresse¹, C. Faure¹, Y. Brans², A. Marais^{1*},

¹ *INRAE, Univ. Bordeaux, UMR BFP, 33140 Villenave d'Ornon, France*

² *Laboratoire de Virologie et de Biologie moléculaire, CTIFL, 24130 Prignonrieux, France*

* Corresponding author: armelle.marais-colombel@inrae.fr

Current address : INRAE, ISVV, Univ. Bordeaux, UMR EGFV, 33140 Villenave d'Ornon Cedex

ORCID ID: 0000-003-2482-1543

Abstract: 99 words

Text (incl. references & figures legends): 2,098 words / 16,183 characters

Figures and Tables: 2 figures and 1 table; 1 supplementary table

The nucleotide sequences reported in this manuscript have been deposited in GenBank under accession numbers OR141115-OR141118, BK064709-BK064710

Keywords: *Prunus*, high-throughput sequencing, *Nepovirus*, *Secoviridae*, ALRSV, MLRSV, stone fruit virus, *Prunus mira*

Abstract

The complete genome sequence of two poorly known *Prunus*-infecting nepoviruses, apricot latent ringspot virus (ALRSV) and myrobalan latent ringspot virus (MLRSV) was determined, confirming they are subgroup C members. Serological, biological and molecular data, in particular a low, 58.8% amino acid coat protein identity suggest ALRSV and MLRSV should be considered distinct. In addition, datamining of wild and ornamental *Prunus* public RNASeq data identified two contigs representing the nearly complete genome of a new subgroup A *Nepovirus* from a smooth stone peach (*Prunus mira*) dataset (SRR8369794) from the Himalayas for which the name *Prunus mira virus A* is proposed.

Members of the *Nepovirus* genus in the family *Secoviridae* (subfamily *Comovirinae*) have bipartite single-stranded, positive-sense RNA genomes encapsidated in icosahedral particles involving a single, large coat protein (CP) subunit. Nepoviruses are divided into three subgroups, A, B, and C based on their genome size and organization, and on their polyprotein cleavage sites [1]. Myrobalan latent ringspot virus (MLRSV) was described in France from a *Prunus cerasifera* tree (myrobalan) showing very poor growth and reduced vigor [2]. While having a known geographical distribution limited to a small area of southwest France, MLRSV was extensively studied in the 1970's and early 1980's [3-5]. This allowed to identify the association of some isolates with a satellite RNA [3] and to conclude that despite initial reports of serological cross-reactions with tomato black ring virus [5], its affinities were with subgroup C nepoviruses characterized by a large RNA2 molecule [4]. Apricot latent ringspot virus (ALRSV) was initially identified during an extensive survey in 1994 in southwest France. It was isolated from apricot trees in commercial apricot orchards showing reduced foliage and a bare appearance but no symptoms on fruits [6]. Transmission to woody or herbaceous hosts allowed its partial characterization and, in particular, the demonstration that it was not serologically related to any tested nepovirus, including MLRSV, leading to the suggestion to consider it as a distinct subgroup C member. Whereas significant serological and biological data is available for these two viruses [2-6], only a very partial genomic RNA2 sequence is available for ALRSV (2,218 nt, AJ278875) [6], while no genomic sequence information is available for MLRSV.

ALRSV and MLRSV isolates are maintained by grafting on GF305 peach seedling indicators and on trees in the virus resource conservatory. Leaves from infected plants were used for total RNAs extraction [7]. RNAs were converted to cDNA, and sequenced (HiSeq3000, 2x150nt paired reads; *ca.* 30 million and 49 million reads, respectively). Following quality trimming, reads were

assembled using CLC Genomics Workbench 21.0.3 (CLC GW, Qiagen), and contigs annotated by Blastn and Blastx against GenBank. In this way, long contigs representing near complete genomic RNAs sequences for ALRSV and MLRSV were identified (Supplementary Table 1). In parallel, a small-scale datamining effort was performed as described above on publicly available RNASeq Sequencing Read Archives (SRAs) from wild or ornamental *Prunus* species. One SRA (SRR8369794, *ca.* 48 million paired reads) derived from smooth stone peach (*Prunus mira*) from the western Himalayas was identified, as yielding two long nepovirus-like contigs (BlastX e-value: 0.0). These contigs were validated and extended by repeated rounds of mapping of residual reads using CLC GW (Supplementary Table 1).

The large contigs representing the two genomic RNAs of ALRSV and MLRSV missed only short terminal sequences. Both 5' and 3' genome ends were determined for each molecule by RACE experiments as described [7]. Full-length RNA1 were determined to be 7,690 nt (OR141117) and 7,678 nt (OR141115), excluding the polyA tail, for ALRSV and MLRSV respectively, while their RNA2 were respectively 5,686 nt (OR141118) and 5,736 nt (OR141116). The sequence determined here for ALRSV shows 99.9% nt identity with the previously available partial RNA2 sequence (AJ278875) and 100% aa identity in the encoded protein.

Since they had been obtained by datamining, no efforts could be made to obtain the extreme ends of the two long contigs representing near a complete genome (RNA1: 7,851 nt, BK064709; RNA2: 3,814 nt, BK064710) reconstructed from the SRR8369794 *P. mira* SRA data. However, these contigs appear to represent the complete genome coding potential. They showed closest affinities to the nepovirus raspberry ringspot virus (RpRSV) but are sufficiently distinct (see below) to be considered as representing a new virus, tentatively named *prunus mira* virus A (PMVA).

The size of 5' untranslated regions (UTRs) was determined to be 43 nt and 73 nt in RNA1 and RNA2 of ALRSV, and 34 nt and 81 nt in RNA1 and RNA2 of MLRSV, respectively. In PMVA, the 5' UTRs are longer, with 126 nt and 173 nt in RNA 1 and RNA2, respectively. The size of the 3' UTRs of both genomic molecules in ALRSV and MLRSV are quite similar (1,233 nt and 1,257 nt in RNA1 and RNA2 of ALRSV, and 1,236 nt and 1,278 nt in RNA1 and RNA2 of MLRSV). The 3' UTRs of MLRSV RNA1 and RNA2 also share 97%, while the 3' UTRs of ALRSV RNA1 and RNA2 are 99.8% identical. On the other hand, these 3' UTRs are shorter in the contigs for PMVA (597 nt and 320 nt in RNA1 and RNA2, respectively), with the last 3' 320 nt of each RNA segment being nearly identical (97.5%).

The three genomes have a typical nepovirus organization [1] with a unique long open reading frame (ORF) encoded on each genomic RNA (Figure 1). As expected, the putative P1 polyproteins encoded by the RNA1 (2,138, 2,136 and 2,376 aa for ALRSV, MLRSV and PMVA, respectively) harbor three conserved domains: a helicase domain (pfam00910), an RNA-dependent RNA polymerase domain (RdRp, cd01699) and, between the Hel and RdRp domains, a conserved Picornavirales 3C/3C-like protease domain (PS51874) identified by ProSite searches. The putative RNA2-encoded P2 polyproteins are 1,452, 1,459 and 1,107 aa-long, respectively and contain three conserved coat protein (CP) domains (pfam03689, pfam03391 and pfam03688) identified by InterProScan (<http://www.ebi.ac.uk/interpro/search/sequence/>) (Figure 1). Sequence comparisons allowed to propose that the P1 polyproteins are probably cleaved into six proteins, as described in the genus, and also allowed a tentative identification of the various cleavage sites [9] (Figure 1). However, while all cleavage sites for ALRSV and MLRSV could be confidently predicted, in the case of PMVA only the three cleavage sites in the central part of the P1 polyprotein could be confidently predicted, while for the two sites towards the polyprotein N-terminal end, the

prediction was not straightforward. The two cleavage sites in the P2 polyprotein liberating the three mature proteins 2A, movement protein (MP) and coat protein (CP) could be confidently identified for all three viruses (Figure 1).

A neighbor-joining tree based on the alignment of the CPs of various *Nepovirus* genus members shows that ALRSV and MLRSV cluster together in the subgroup C, while PMVA is grouped with subgroup A members and is most closely related to RpRSV (Figure 2). Indeed, sequence comparisons showed that the P1 and P2 polyproteins of PMVA have respectively 64% and 63% aa identity with those of RpRSV. These identity levels are clearly below the currently accepted molecular species cut-off values in the *Nepovirus* genus (75% aa identity for CP and 80% aa identity for Pro-Pol region [1]). A similar phylogenetic analysis performed using the Pro-Pol region provided the same clustering of ALRSV, MLRSV and PMVA (data not shown), but with ALRSV and MLRSV much more closely related. Indeed, ALRSV and MLRSV share only 58.8% aa identity in their CP (below the 75% species cut-off) but 94% aa identity in the taxonomically informative Pro-Pol region of the P1 polyprotein (above the 80% species cut-off) [1].

The genomic characterization of ALRSV and MLRSV provided here confirms their identification as nepoviruses based on the biological and partial molecular characterization previously available [2-6]. In particular, phylogenetic analyses together with the long highly conserved 3' UTRs between the two genomic RNAs clearly indicate that ALRSV and MLRSV belong to the genus subgroup C, in contrast to PMVA which belongs to subgroup A. [1]. Unlike a previous report that described a satellite RNA associated with MLRSV [3, 4], we did not identify any satellite RNA associated with either the MLRSV or the ALRSV isolates analyzed here. Due to the fact that only one region (CP) identity value meets the species demarcation criteria [1], the taxonomic position of ALRSV and MLRSV cannot be straightforwardly decided. Both viruses also share some

common biological properties: they were both identified in the south of France, and display similarities in symptomatology when grafted on the same woody host, such as symptoms of leaf enation and russetting on sweet cherry (cv. Bing), leaf enation and chlorotic spots on leaves, shortened internodes and stunting on GF305 peach seedlings [6]. In addition, when ALRSV was grafted on *P. cerasifera*, the natural host of MLRSV, symptoms similar to those induced by MLRSV were observed, in particular shortened internodes [6]. However, some differences were observed [2-6] following experimental inoculation of herbaceous hosts: while MLRSV caused a systemic mosaic on *Cucumis sativus*, ALRSV induced no symptoms, and the reverse was true on *Nicotiana clevelandii*, with ALRSV inducing a systemic necrotic mosaic but MLRSV causing an asymptomatic infection. In addition, no serological relationships were identified between ALRSV and any of the tested fruit tree-infecting nepoviruses, including MLRSV [6].

An analysis of recombination events performed using multiple alignments of genomic RNA1 or RNA2 of nepoviruses using the RDP4 program [8] with default parameters failed to detect any significant events involving the RNA1 or RNA2 molecules of MLRSV or ALRSV. In addition, a detailed pairwise comparison was performed to determine whether a reassortment of genomic segments might be responsible for the large differences in identity levels observed between the RNA1-encoded Pro-Pol region and the RNA2-encoded CP of MLRSV and ALRSV. This analysis demonstrated that the RNA1 and RNA2 of these two viruses share 80.4% and 74.1% nt identity, respectively (Table 2). The corresponding values for the 5' UTRs are 55-55.9% and 86.3-86.5% for the 3' UTR, respectively. Looking at the encoded proteins, the region of the RNA1-encoded polyprotein upstream of the Pol-Pro region (corresponding to the X1, X2, NTB and Vpg) share 77.3% aa identity between MLRSV and ALRSV, while the RNA2-encoded 2A and MP share respectively 65.4% and 88.4% aa identity. Rather than a genomic segment reassortment or a

recombination event, it thus appears that the very different identity levels observed between the Pro-Pol and CP regions of MLRSV and ALRSV are more likely the result of wide differences in sequence conservation between different parts of the viral polyproteins. These may in turn indicate that particular selection pressures may apply to some viral proteins (Pro-Pol, MP), possibly as a consequence of host adaptation.

Altogether, considering the genomic divergence between the two viruses, the lack of serological relationships and the differences observed in symptomatology in some herbaceous hosts, and even if the Pro-Pol based species discrimination criterion is not met [1], we propose that ALRSV and MLRSV should be considered as two distinct species. The availability of complete genomic sequences for these two *Prunus*-infecting nepoviruses should be useful for the development of specific diagnostic assays in support of certification programs and of further studies of the epidemiology of these still poorly known viruses. In the case of PMVA, given that it was only identified by datamining and that its presence in *P. mira* could not be experimentally confirmed, its status as a novel *Prunus*-infecting nepovirus remains to be confirmed but the availability of a near complete genomic sequence opens the way to the further characterization of this novel virus and, in particular, to the determination of its geographic distribution, host range and potential pathogenicity.

Acknowledgements

The authors thank the GeT-PlaGe Platform (GenoToul, INRAE, Toulouse, France) for high-throughput sequencing and François Latour for taking care of the virus resource conservatory at CTIFL.

References

1. Thompson JR, Dasgupta I, Fuchs M, Iwanami T, Karasev AV, Petrzik K, Sanfaçon H, Tzanetakis I, van der Vlugt R, Wetzel T, Yoshikawa N, ICTV Report Consortium (2017) ICTV Virus Taxonomy Profile: *Secoviridae*. J Gen Virol 98:529-531.
2. Dunez J, Delbos R, Desvignes JC, Marenaud C, Kuszala J, Vuittenez A (1971) Evidence of a ringspot type virus on *Prunus cerasifera* (myrobalan B). Annales Phytopath HS:117-128.
3. Fritsch C, Koenig I, Murrant AF, Raschke JH, Mayo MA (1984) Comparisons among Satellite RNA species from five isolates of tomato black ring virus and one isolate of myrobalan latent ringspot virus. J Gen Virol 65:289-294.
4. Gallitelli D, Piazzolla P, Savino V, Quacquarelli A, Martelli GP (1981). A comparison of myrobalan latent ringspot virus with other nepoviruses. J Gen Virol 53:57-65.
5. Dunez J, Delbos, R (1976) Myrobalan latent ringspot virus, a bipartite genome virus and a strain of tomato black ring virus. Acta Hort 67:324-324.
6. Gentit P, Delbos RP, Candresse T, Dunez J (2001) Characterization of a new nepovirus infecting apricot in Southeastern France: apricot latent ringspot virus. Eur J Plant Pathol 107:485-494.
7. Khalili M, Candresse T, Koloniuk I, Safarova D, Brans Y, Faure C, Delmas M, Massart S, Aranda MA, Cagalayan K, Decroocq V, Drogoudi P, Glasa M, Pantelidis G, Navratil M, Latour F, Spak J, Pribylova J, Mihalik D, Palmisano F, Saponari A, Necas T, Sedlak J, Marais A (2023) The expanding menagerie of *Prunus*-infecting luteoviruses. Phytopathology 113:345-354.

8. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B (2015) RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus evolution*, 1:vev003.
9. Sanfaçon H (2022) Re-examination of nepovirus polyprotein cleavage sites highlights the diverse specificities and evolutionary relationships of nepovirus 3C-like proteases. *Arch Virol* 167:2529-2543.

Compliance with Ethical Standards:

Funding: This study was funded by the European Union through the Horizon 2020 Marie Skłodowska-Curie Actions Innovative Training Network (H2020 MSCA- 60 ITN) project “INEXTVIR” (Grant agreement number 813542).

Conflict of Interest: M. Khalili declares that she has no conflict of interest. T. Candresse declares that he has no conflict of interest. C. Faure declares that she has no conflict of interest. Y. Brans declares that he has no conflict of interest. A. Marais declares that she has no conflict of interest.

Ethical approval : This article does not contain any studies with human participants or animals performed by any of the authors.

Statements and Declarations

Author Contributions: All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Maryam Khalili, Thierry Candresse, Chantal Faure, Yoann Brans and Armelle Marais. The first draft of the manuscript was written by Maryam Khalili and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.”

Data Availability : The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.”

Figure legends

Fig 1. Genome organization of the three *Prunus*-infecting nepoviruses apricot latent ringspot virus (ALRSV) (**A**), mirobalan latent ringspot virus (MLRSV) (**B**), prunus mira virus A (PMVA) (**C**). Open reading frames are indicated by boxes with the different viral proteins in different shades of grey separated by dashed lines marking predicted cleavage sites. The predicted dipeptides at cleavage sites and their position on viral polyproteins are shown by black triangles (? indicates tentative cleavage sites). Nucleotide positions at 3' and 5' ends are marked on the line representing untranslated regions. Identified conserved protein motifs [Helicase, RNA-dependent RNA polymerase (RdRp), and coat protein (CP)] are shown as circles within proteins. 2A: protease cofactor; NTB: nucleotide triphosphate binding helicase; VPg: viral genome-linked protein; Pro: protease; Pol: polymerase; MP: movement protein; X1 and X2: putative X1 and X2 proteins; A(n) polyA tail.

Fig. 2. Neighbour-joining phylogenetic tree based on the alignment of the coat protein sequences of various *Nepovirus* members. The tree was reconstructed using strict amino acid identity p-distances (randomly bootstrapping 1,000 replicates). Bootstrap values less than 70% were removed. Accession numbers of reference sequences in GenBank are given, followed by the virus name. The viruses characterized in this study are shown by black circles. The scale bar represents 10% aa divergence.

Table 1. Pairwise comparison between ALRSV and MLRSV genomic regions

Genome segment	Full genome nt identity	Polyprotein aa identity	5' UTR nt identity	3' UTR nt identity
RNA1	80.4%	84.0%	55.9%	86.3%
RNA2	74.1%	69,7%	55%	86.5%

Figure 1

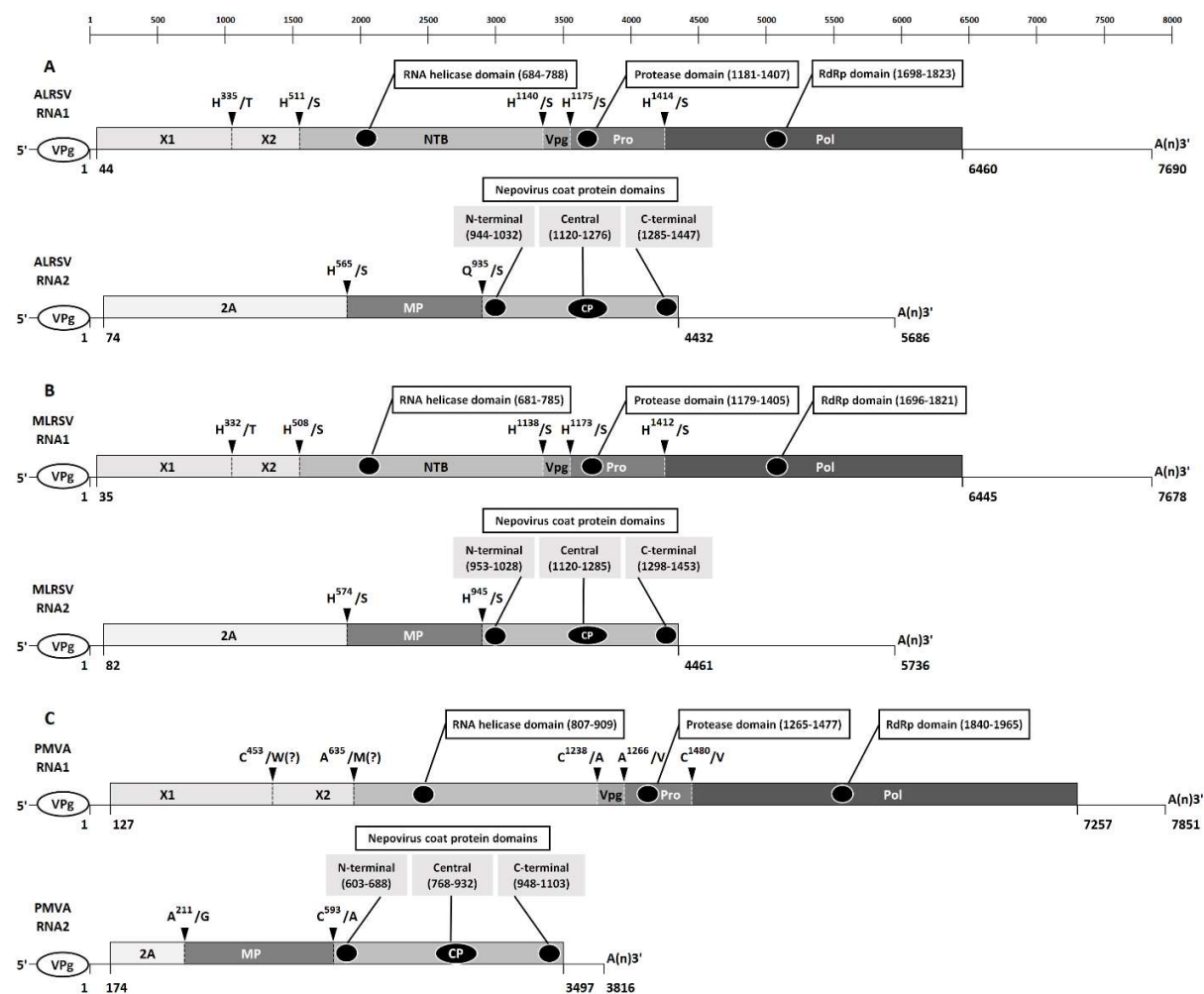
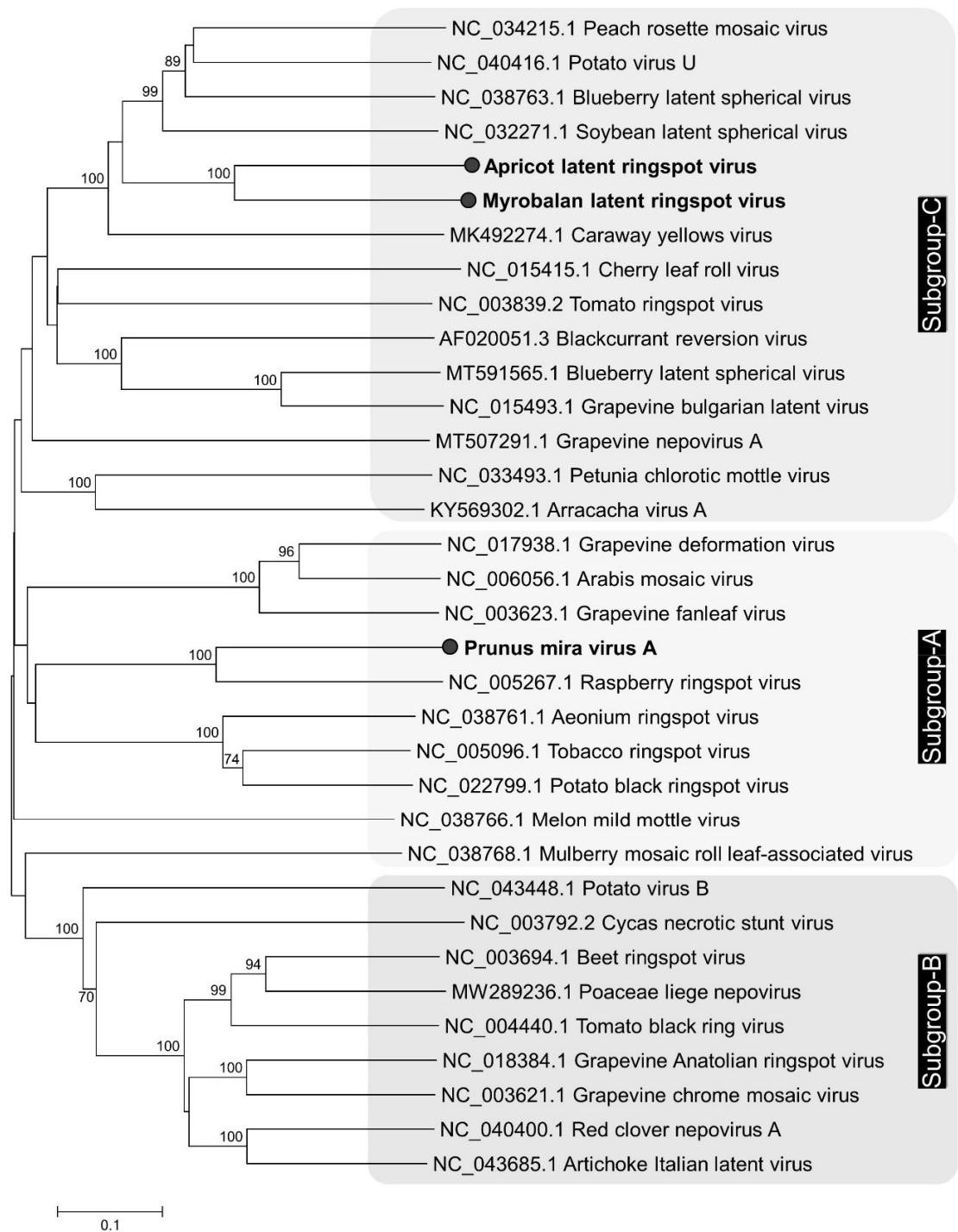


Figure 2



Supplementary Table 1. Number of trimmed reads, average coverage and percentage of total reads mapped on viral genome segments of ALRSV, MLRSV and PMVA

Virus	Segment	Number of mapped reads	Percentage of mapped reads	Average coverage
ALRSV	RNA1	157,738	0.3%	3,091x
	RNA2	235,363	0.48%	6,080x
MLRSV	RNA1	62,905	0.21%	1,416x
	RNA2	121,600	0.4%	1,883x
PMVA	RNA1	7,036	0.01%	132x
	RNA2	3,825	0.007%	146x