Complete genome sequence of almond luteovirus 1, a novel luteovirus species infecting almond

Maryam Khalili, Thierry Candresse, Chantal Faure, Armelle Marais *

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

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

ORCID ID A. Marais: 0000-0003-2482-1543

ORCID ID T. Candresse: 0000-0001-9757-1835

Abstract: 136 words

Text (incl. references & figures legends): 8432 characters including spaces

Figures and Tables: 2 figures and 1 supplementary figure

The nucleotide sequence reported in this manuscript has been deposited in GenBank under
accession number MT362517

Keywords: stone fruits, virome, high-throughput sequencing, *Luteoviridae*, *Prunus*

Abstract

In this study, we report the complete genome sequence of a novel luteovirus detected in almonds using high-throughput sequencing. The genome of the new luteovirus comprises 5,047 nucleotides, and its genomic organization is similar to that of the recently described nectarine stem pitting associated virus (NSPaV), with only four open reading frames encoding for replication related proteins, the coat protein (CP) and a CP readthrough protein involved in the aphid transmission of luteovirids. Phylogenic and pairwise distance analyses showed that this virus shares 79% and 57.8% amino acid identities in the P1-P2 fusion protein and the P3-P5 protein, respectively, with the most closely related luteovirus NSPaV, suggesting that it represents a novel species for which the name of *Almond associated luteovirus 1* is proposed. To our knowledge, this is the first report of an almond-infecting luteovirus.

INTRODUCTION

Almond (*Prunus dulcis*) is believed to have originated from Southwest Asia where it was initially brought into domestication and is today cultivated in many countries. Graft-transmissible pathogens (viruses, viroids and phytoplasmas) cause economically important diseases in almond, in particular viruses of the genera *Ilarvirus*, *Trichovirus*, *Nepovirus* and *Potyvirus* [1]. However, in contrast to several other *Prunus* species, no luteovirus has thus far been reported from almond.

Luteoviruses belong to the family *Luteoviridae*, which also comprises the genera *Enamovirus* and *Polerovirus* [2]. Three luteoviruses, nectarine stem pitting-associated virus (NSPaV) [3], cherry-associated luteovirus (ChALV) [4] and peach-associated luteovirus (PaLV) [5] have recently been discovered in *Prunus*. The genome of luteoviruses varies between 5 and 5.9 kilobases (kb) [2]. Among luteoviruses described so far, NSPaV is the sole member with only four open reading frames (ORFs) [3], whereas other members contain six or seven ORFs [2].

As part of a virus discovery effort, double-stranded RNAs were purified from the leaves of an almond tree without obvious viral symptoms collected in Turkey in the spring of 2019. Following reverse transcription and random amplification [6], the obtained cDNAs were sequenced (HiSeq3000 Illumina platform, 2x150 bp). After reads cleaning, assembly into contigs and Blast-based annotation [7], a single 5 kb contig showing significant nucleotide (nt) identity with luteoviruses was identified. This contig integrated 35,411 reads (34.3% of total reads) for an average coverage of 782x and showed 71% nt identity with the NSPaV genome (NC027211) which spanned from nt 39 to 4,991. The genome sequence was completed by determining the genome ends using the SMARTer RACE Kit (Takara) and has been deposited in GenBank under accession number MT362517.

The complete genome of the almond luteovirus is 5,047 nt long and comprises four ORFs, homologous to the corresponding NSPaV ORFs (Fig. 1). ORF1 encodes the putative RNA-dependent RNA polymerase (RdRp) of 329 amino acids (aa), while an 848 aa fusion protein is expressed by a -1 ribosomal frameshift between ORF1 and ORF2. ORF3 encodes the

putative 204 aa coat protein (CP) and ORF5 is expressed as a readthrough fusion to the CP (550 aa) following the suppression of the CP stop codon (Fig. 1). With 637 nt, the long 3' untranslated region (UTR) has a length similar to that of NSPaV, while the 5' UTR is 149 nt long. Similar to NSPaV, ORF5 (345 aa) was found to be significantly shorter than in other luteoviruses, and neither ORF3a [8] nor ORF4 (putative movement protein) nor ORF6 could be identified by scanning the genome and analyzing translation in all reading frames.

A phylogenetic analysis using complete luteovirus genomes (GenBank) confirms that the almond virus belongs to the genus *Luteovirus* and has the closest affinities with NSPaV with which it shares 71.5% nt identity over the whole genome (Supplementary Fig. 1). Phylogenetic trees reconstructed using amino acid sequences of the RdRp or CP showed that NSPaV and the almond virus form a specific branch, the RdRp tree showing a lower branching as compared to the CP tree (Fig. 2A and B). Pairwise comparisons for the various viral proteins showed that the almond virus P1-P2 fusion protein shares 49.6%-79% aa identity with those of luteoviruses, the highest being NSPaV. For the CP-P5 readthrough protein, the level of identity was even lower (27.2%-57.8%). Given the currently accepted molecular species demarcation criteria for luteoviruses (more than 10% in aa divergence in any gene product [2]), the almond virus defines a novel *Luteovirus* species for which the name *Almond luteovirus* 1 (AILV1) is proposed. In order to clarify potential recombination events with related viruses, a recombination analysis was performed using RDP 4 on a multiple alignment of representative isolates of recognized luteovirus species but failed to identify statistically supported recombination events involving AILV1.

To estimate the incidence of AILV1 in *Prunus*, detection primers were designed (NLuteo-PoIF1 5'-GCACATGTTTCGAGGAGATG-3', NLuteo-PoIR1 5'-CCGTCCTCAGATCTTTCCAT-3') targeting a fragment of 505 nt in the polymerase gene. Total nucleic acids were extracted [9] from 105 *Prunus* samples (*P. dulcis*, *P. sibirica*, *P. avium*, *P. armeniaca*, *P. cerasifera*, *P. domestica*, *P. mume* and *P. persica*) collected from Azerbaijan, Kazakhstan, China, Italy and France and submitted to a two-step RT-PCR [10]. Besides the positive sample, none of the tested samples were found to be infected by AILV1. To our best knowledge, AILV1 is the first

92 described luteovirus infecting almonds. Further studies are required to estimate its distribution
93 in stone fruits, as well as its potential pathogenicity. As for NSPaV, the absence of ORFs 3a,
94 4 and 6 and the presence of a shorter readthrough domain might suggest that its biology could
95 significantly differ from that of other luteoviruses.

96

Acknowledgements

The authors thank the INRAE GenoToul Platform (Toulouse, France) for HTS, and Eden Darnige for proofreading the manuscript.

Compliance with ethical standards: this manuscript presents original work that does not involve studies with humans or animals.

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

References

1. Savino V, Bazzoni A, Boscia D (2011) Virus and Phytoplasma Diseases of Almond. In Hadidi A, Barba M, Candresse T, Jelkmann W (eds) Virus and virus-like diseases of pome and stone fruits. APS Press, pp 271-276.
2. Domier LL (2012) Family *Luteoviridae*. In: King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (eds) Virus taxonomy: ninth report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, pp 1045-1053.
3. Bag S, Al Rwahnih M, Li A, et al (2015) Detection of a New Luteovirus in Imported Nectarine Trees: A Case Study to Propose Adoption of Metagenomics in Post-Entry Quarantine. *Phytopathology* 105(6):840-846.
4. Lenz O, Pribylova J, Franova J, et al (2016) Identification and characterization of a new member of the genus Luteovirus from cherry. *Arch Virol* 162:587-590.
5. Wu LP, Liu HW, Bateman M, Liu Z, Li R (2017) Molecular characterization of a novel luteovirus from peach identified by high-throughput sequencing. *Arch Virol* 162:2903-2905.
6. Marais A, Faure C, Bergey B, et al (2018) Viral Double-Stranded RNAs (dsRNAs) from Plants: Alternative Nucleic Acid Substrates for High-Throughput Sequencing. *Methods Mol Biol* 1746:45-53.
7. Candresse T, Marais A, Faure C, et al (2013) Association of *Little cherry virus 1* (LChV1) with the Shirofungen stunt disease and characterization of the genome of a divergent LChV1 isolate. *Phytopathology* 103(3):293-298.

- 124 8. Smirnova E, Firth AE, Miller WA et al (2015) Discovery of a Small Non-AUG-Initiated ORF
125 in Poleroviruses and Luteoviruses That Is Required for Long-Distance Movement. PLoS
126 Pathog 11(5): e1004868.
- 127 9. Foissac X, Svanella-Dumas L, Gentit P, et al (2005) Polyvalent degenerate oligonucleotides
128 reverse transcription-polymerase chain reaction: a polyvalent detection and
129 characterization tool for trichoviruses, capilloviruses, and foveaviruses.
- 130 10. Marais A, Svanella-Dumas L, Barone M, et al (2012) Development of a polyvalent RT-
131 PCR detection assay covering the genetic diversity of Cherry capillovirus A. Plant Pathol
132 61:195-204.
- 133

Figure1. Schematic representation of the genome organization of almond luteovirus 1.

Boxes represent the Open Reading Frames (ORF) and numbers indicate the coordinates of each ORF. A dashed line shows the position of stop codon between the coat protein (CP) and readthrough protein. RT: readthrough; RTD: readthrough domain; -1FSh: -1 ribosomal frameshift; RdRp: RNA-dependent RNA polymerase

141

142

Figure 2. Neighbor-joining phylogenetic trees reconstructed using the alignments of the amino acid sequences of the RNA-dependent RNA polymerase (A) and of the coat protein (B) of luteoviruses with the type polerovirus (potato leaf roll virus) and enamovirus (pea enation mosaic virus-1) used as outgroups. Only bootstrap values above 70% are shown (1,000 replicates). Almond luteovirus 1 is indicated by the black dot. Accession numbers are given. The scale bars represent 10% (A) and 5% (B) amino acid divergence.

Supplementary Figure 1. Neighbor-joining phylogenetic tree based on the complete genome sequences of luteoviruses with potato leaf roll virus and pea enation mosaic virus-1 used as outgroups. Only bootstrap values above 70% are shown (1,000 replicates). Almond luteovirus 1 is indicated by the black dot. Accession numbers are given in parentheses. The genera to which the various viruses belong are indicated on the right. The scale bar represents 5% nucleotide divergence.