Scanning the virome of Prunus : Identification and characterization of novel luteoviruses and Secoviridae members

Maryam Khalili

▶ **To cite this version:**

Maryam Khalili. Scanning the virome of Prunus : Identification and characterization of novel luteoviruses and Secoviridae members. Vegetal Biology. Université de Bordeaux, 2022. English. NNT : 2022BORD0281 . tel-04777000

HAL Id: tel-04777000

https://theses.hal.science/tel-04777000v1

Submitted on 12 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE PRÉSENTÉE

POUR OBTENIR LE GRADE DE

DOCTEUR DE

L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE SCIENCE DE LA VIE SE DE LA SANTE

SPECIALITE MENTION BIOLOGIE VEGETALE

Par Maryam KHALILI

SCANNING THE VIROME OF PRUNUS: IDENTIFICATION AND CHARACTERIZATION OF NOVEL LUTEOVIRUSES AND SECOVIRIDAE MEMBERS

Sous la direction de : Armelle MARAIS

Soutenue le 26 octobre 2022

Membres du jury :

Mme RAVNIKAR, Maja	Professeure, NIB, Slovénie	Rapportrice
M. JELKMANN, Wilhelm	Professeur, JKI, Allemagne	Rapporteur
M. LEMAIRE, Olivier	Directeur de Recherche,INRAE, Colmar	Examinateur
Mme FORGET, Florence	Directrice de Recherche, INRAE, Bordeaux	Président
M. GENTIT, Pascal	Ingénieur de Recherche, ANSES, Angers	Examinateur
Mme MARAIS Armelle	Ingénieure de Recherche, INRAE, Bordeaux	Directrice de thèse

SCANNING DU VIROME DE PRUNUS : IDENTIFICATION ET CARACTERISATION DE NOUVEAUX LUTEOVIRUS ET DE NOUVEAUX MEMBRES DES SECOVIRIDAE

Résumé :

Les avancées majeures en biologie ont souvent été liées à des innovations technologiques importantes. Les technologies de séquençage haut débit (HTS, High-Throughput Sequencing) sont sans aucun doute l'une des inventions les plus importantes de ces dernières années, impactant entre autres et de façon majeure la découverte de nouveaux virus. Les HTS permettent de caractériser sans *a priori* le virome d'un échantillon, soit l'ensemble des virus présents. Néanmoins, l'avènement des HTS n'a été possible que grâce aux progrès des outils d'analyse bio-informatique permettant d'exploiter les données de séquençage massif. La présente Thèse a été réalisée dans le cadre d'un réseau de formation Marie-Curie nommé « INEXTVIR – Innovative Network for Next Generation Training and Sequencing of Virome » qui a pour but de mieux comprendre les communautés virales et leur rôle dans les écosystèmes agricoles, en utilisant les dernières avancées en matière de HTS. L'objectif central de cette thèse a été de déterminer la composition du virome de plusieurs espèces de *Prunus* afin d'identifier les virus, connus ou nouveaux, qui le composent, de caractériser certains d'entre eux et de développer des outils de diagnostic. Afin de maximiser la diversité génétique au sein des espèces de Prunus ciblées, le virome de 300 accessions provenant d'espèces cultivées ou apparentées, conservées dans le Centre de Ressources Biologiques (CRB) *Prunus* d'INRAE a été analysé. Ainsi, des accessions de *Prunus armeniaca* (abricotier), *P. persica* (pêcher), *P. domestica* (prunier), *P. avium* (cerisier doux), *P. cerasus* (cerisier acide), et d'espèces apparentées ont été soumises à un indexage par HTS. Des Prunus sauvages ou ornementaux collectés au gré d'autres études ont également été analysés. Pour cela, les ARN doubles brins, proxy d'une infection virale par des virus à ARN, ont été purifiés puis analysés. Au-delà des virus classiquement connus pour infecter les espèces du genre *Prunus*, quatre nouveaux virus appartenant au genre *Luteovirus*, ont été mis en évidence chez divers *Prunus* : pêcher, abricotier, cerisier acide, abricotier du Japon (*P. mume*), un cerisier ornemantal (*P. incisa*) ou une espèce sauvage *P. mahaleb*. Par ailleurs, une recherche de séquences virales dans les bases de données Prunus a permis de reconstruire le génome d'un autre lutéovirus infectant potentiellement une espèce sauvage, *Prunus humilis*. Avant

cette étude, seulement trois luteovirus étaient connus pour infecter les *Prunus*. L'identification et la caractérisation de ces cinq nouveaux luteovirus représentent donc une avancée importante dans la connaissance de ces agents. Une étude menée à l'échelle européenne a permis de montrer que les trois luteovirus (dont un nouvel agent) infectant le pêcher étaient très largement répandus avec des taux de prévalence importants, corroborant les données obtenues au sein de la collection de pêchers du CRB Prunus. Des essais de transmission de ces trois luteovirus à l'indicateur biologique GF305, ainsi que l'observation des arbres du CRB infectés suggèrent que ces virus n'induisent très vraisemblablement pas de symptômes significatifs, ce qui expliquerait qu'ils soient passés inaperçus jusqu'à présent, en dépit de leur très large distribution. Par ailleurs, la séquence génomique complète de quatre virus de la famille des *Secoviridae* a été déterminée au cours de ce travail, dont un nouveau *Cheravirus* identifié et caractérisé dans deux espèces sauvages (*P. mahaleb* et *P. brigantina*) et dans une accession d'abricotier. Des tests de détection par RT-PCR performants ont été développés pour chacun des nouveaux virus étudiés. Globalement, ces travaux d'exploration du virome des Prunus ont permis d'enrichir notre connaissance des virus infectant cette famille et d'apporter de nouveaux éléments permettant de commencer à évaluer les risques potentiellement liés à ces différents agents. Ils ouvrent de nouvelles perspectives de recherche pour rendre compte de leur impact.

Mots clés : Nouveau virus, Virome, *Prunus*, High Throughput Sequencing, Métagénomique

SCANNING THE VIROME OF PRUNUS: IDENTIFICATION AND CHARACTERIZATION OF NOVEL LUTEOVIRUSES AND SECOVIRIDAE MEMBERS

Abstract:

Major advances in biology have often been linked to important technological innovations. High-Throughput Sequencing (HTS) technologies are undoubtedly one of the most important inventions of recent years, having a major impact on the discovery of new viruses. HTS allows to characterize without a priori the virome of a sample, *i.e.* all the viruses present. Nevertheless, the advent of HTS was only possible thanks to the concomitant progress of bioinformatics tools allowing to exploit massive sequencing data. The present thesis was carried out in the framework of a Marie-Curie training network named « INEXTVIR - Innovative Network for Next Generation Training and Sequencing of Virome » which aims to better understand viral communities and their role in agricultural ecosystems, using the latest advances in HTS. The main objective of this thesis was to determine the composition of the virome of several *Prunus* species in order to identify known and novel viruses that compose it, to characterize some of them and to develop diagnostic tools, allowing their detection. In order to maximize the genetic diversity within the targeted *Prunus* species, the virome of 300 accessions from cultivated or related species, gathered in the INRAE Prunus Biological Resource Center (BRC) was analyzed. Thus, accessions of *Prunus armeniaca* (apricot), *P. persica* (peach), *P. domestica* (plum), *P. avium* (sweet cherry), *P. cerasus* (sour cherry), and related species were subjected to HTS indexing. Wild or ornamental *Prunus* collected in other studies were also analyzed. For this purpose, double-stranded RNAs, proxy of viral infection by RNA viruses, were purified and analyzed. In addition to the viruses classically known to infect *Prunus* species four new viruses belonging to the genus *Luteovirus* were identified in various *Prunus* species: peach, apricot, sour cherry, Japanese apricot (*P. mume*), an ornamental cherry (*P. incisa*) or a wild species, *P. mahaleb*. In addition, a search in the *Prunus* databases for viral sequences allowed the reconstruction of the genome of another luteovirus potentially infecting a wild species, *Prunus humilis*. Prior to this study, only three luteoviruses were known to infect *Prunus* trees. The identification and characterization of these five new luteoviruses therefore represent an important advance in the knowledge of these agents. A Europe-wide study showed that the three peach-infecting luteoviruses (including one new agent) were

widely distributed with high prevalence rates, in agreement with the data obtained in the peach collection of the *Prunus* BRC. Transmission assays of these three luteoviruses to the biological indicator GF305, as well as observation of infected trees in the BRC, suggest that these viruses most likely do not induce significant symptoms, which would explain why they have gone undetected until now, despite their very wide distribution. In addition, the complete genomic sequence of four viruses of the *Secoviridae* family was determined during this work, including a new cheravirus identified and characterized in two wild species (*P. mahaleb* and *P. brigantina*) and in an apricot accession from the BRC. Efficient RT-PCR detection tests were developed for each of the new viruses studied. Overall, this exploration of the *Prunus* virome has enriched our knowledge of the viruses infecting this family and has provided new elements to start assessing the potential risks linked to these different agents. These results open new research perspectives to account for their impact.

Keywords: Novel virus, Virome, *Prunus*, High Throughput Sequencing, Metagenomics

---

Unité de recherche

UMR 1332 Biologie du Fruit et Pathologie, INRAE and Université de Bordeaux, Campus INRAE de la Grande Ferrade, 71 avenue Edouard Bourleaux, CS20032, 33882 Villenave d'Ornon

ACKNOWLEDGEMENTS

I am extremely grateful to my supervisor, **Dr. Armelle Marais**, for her invaluable guidance, ongoing support, and patience throughout my PhD. Her immense knowledge and plentiful experience have encouraged me all the time in my academic research and daily life. I want to thank her for her support, both technically and emotionally. I would not have been able to cope with many challenges without her tremendous understanding and kind helps. I thank her for being an adorable, meticulous supervisor who keeps the laboratory running smoothly, and for her dedication and concern for the well-being of the team members.

I would like to express my deepest appreciation to my co-supervisor **Dr. Thierry Candresse** for his knowledge and expertise he shared with me. This endeavor would not have been possible without his inspiration, advice, suggestions and feedback. I was fortunate enough to work in an office in front of his, so I often went there to ask my questions and he was always generously offering advice and encouragement with a perfect blend of insight and humor.

I'd like to acknowledge the thesis committee members **Dr. Maja Ravnikar**, **Dr. Wilhelm Jelkmann** for the evaluation of my thesis and the opportunity to benefit from their knowledge, advice and feedback and **Dr. Florence Forget**, **Dr. Pascal Gentit** and **Dr. Olivier Lemaire** for taking advantage of their expertise and insight.

I would also like to extend my thanks to my mid-term committee members **Dr. Miroslav Glasa**, **Dr. Antonio Olmos** and my tutor **Dr. Macha Nikolski** for their encouraging words and thoughtful, detailed feedback and advice, and **Dr. Sébastien Massart**, not only for his invaluable advice and support but also for hosting me at the University of Liége along with his team **Nuria**, **Johan** and **Coline,** during my secondment. I am so thankful for the opportunity of working with such amazing people.

I would like to highlight one truly exceptional person from the etiology team **Chantal Faure** who was always there with me to offer both practical and emotional support. Her company in the lab and in my life will be always remembered.

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).

ACKNOWLEDGEMENT

I am also grateful to INEXTVIR project coordinators and managers, particularly **Dr. Neil Boonham** for his constant support and arranging both specific and multidisciplinary training over the last three years.

Special thanks to our dear advisor within INEXTVIR project **Dr. Carolyn Malmstrom** who gave us the opportunity of discussion and criticization of the scientific papers through the online journal club holding every two weeks. Many thanks to her for the infinitive support and encouragement.

I had the pleasure of working and collaborating with other ESRs in the INEXTVIR project, more closely with **Deborah**, **Grigorii**, **Bisola**, **Paul**, and **Ayoub**.

I must thank all the colleagues in the etiology team, **Marie** and **Laurence** as well as the entire virology group of INRAe for making the past 3 years much more enjoyable and keeping me encouraged throughout the whole process.

I would also like to thank my son **Taha**, who missed me a lot during my PhD, but I was heartened by his comprehension.

Finally, I owe my deepest gratitude to **Morteza**, who is my love, my best friend, and my husband. I am forever grateful for unconditional love and support all along the thesis process and every day. I am deeply indebted to him for leaving his life behind so that my dreams come true, and I would like to dedicate this thesis to him.

SCIENTIFIC PRODUCTIONS

- *List of publications*

  - **Khalili, M**., Candresse, T., Faure, C., & Marais, A. (2020). Complete genome sequence of almond luteovirus 1, a novel luteovirus infecting almond. *Archives of Virology*, *165*(9), 2123-2126. https://doi.org/10.1007/s00705-020-04715-2
  - **Khalili, M**., Candresse, T., Koloniuk, I., Safarova, D., Brans, Y., Faure, C., ... & Marais, A. (2022). The expanding menagerie of Prunus-infecting luteoviruses. *Phytopathology*, (ja). https://doi.org/10.1094/PHYTO-06-22-0203-R
  - Sukhorukov, G., **Khalili, M**., Gascuel, O., Candresse, T., Marais-Colombel, A., & Nikolski, M. (2022). VirHunter: A Deep Learning-Based Method for Detection of Novel RNA Viruses in Plant Sequencing Data. *Frontiers in Bioinformatics*, 38. https://doi.org/10.3389/fbinf.2022.867111
  - **Khalili, M**., Candresse, T., Brans, Y., Faure, C., Audergon, JM., Decroocq, V., Roch, G., & Marais, A. (2022) Molecular characterization of a new *Prunus*-infecting cheravirus and complete genome sequence of stocky prune virus. *Viruses.* 14(11), 2325. https://doi.org/10.3390/v14112325
  - **Khalili, M**., Candresse, T., Brans, Y., Faure, C., & Marais, A. (2022) Complete genome sequences of three *Prunus*-infecting nepoviruses: apricot latent ringspot virus, myrobolan latent ringspot virus and a novel nepovirus from smooth stone peach (*Prunus mira* Koehne). Viruses. To be submitted.

- *List of Oral Communications*

  - September 2021, 18émes Rencontres de Virologie Végétale, Aussois, France
  - June 2022, 11es journées des doctorants du département (SPE) de I'INRAE, Bordeaux, France
  - October 2022, The AAB-organized "International Advances in Plant Virology (IAPV)" conference, Ljubljana, Slovenia

- *Awards*

  - February 2021, Merit Scholarship in plant pathology, in the memory of "Monique Garnier-Semancik"

Catégorie : 2. Mener à bien ses activités de recherche

- Autumn-Spring online seminar series+ school (12 november 2020) organized by INEXTVIR project (no.813542) Marie Sklodowska-Curie Inovative Training Network (MSCN-ITN), 20 heures enregistrées par : Sciences de la Vie et de la Santé.
- Éthique de la recherche (18 mai 2021) Université de Lyon, 12 heures enregistrées par : Sciences de la Vie et de la Santé.
- INEXTVIR Training school 1 (21 avril 2020) Natioal institute of Biology (NIB) Slovenia online, 20 heures enregistrées par : Sciences de la Vie et de la Santé.
- Mapping genetic variations to three-dimensional protein structures (14 avril 2021) Iranian phytopatological society, 4 heures enregistrées par : Sciences de la Vie et de la Santé.
- Online training course on Real time PCR (09 mars 2020) Iranian phytopatological society 4 heures enregistrées par : Sciences de la Vie et de la Santé.
- Research integrity in scientific professions (01 juillet 2020) Université de Bordeaux 15 heures enregistrées par : Sciences de la Vie et de la Santé.
- **Catégorie : 3. Communiquer en langue étrangère**
- Ateliers de conversation en langue anglaise Niveau Advanced C1/C2 groupe 2 : (11 février 2020 - 23 avril 2020) Site Talence - DLC bât A21, 24 heures
- Français Langue Etrangère/DOCTOFLE/French courses for PhD (24 novembre 2020) à distance plateforme ZOOM, 34 heures
- Français Langue Etrangère/DOCTOFLE/French courses for PhD (26 novembre 2019) cours Campus Talence - Bât A21, 40 heures
- Français Langue Etrangère/DOCTOFLE/French courses for PhD 2° session (16 mars 2020) cours Campus Talence - Bât A21, 40 heures
- **Catégorie : 4. Développer ses compétences relationnelles**
- Writing better English (19 mai 2020) CABI online-Platform Webinar 2 heures enregistrées par : Sciences de la Vie et de la Santé.
- **Catégorie : 6. Préparer sa poursuite de carrière**

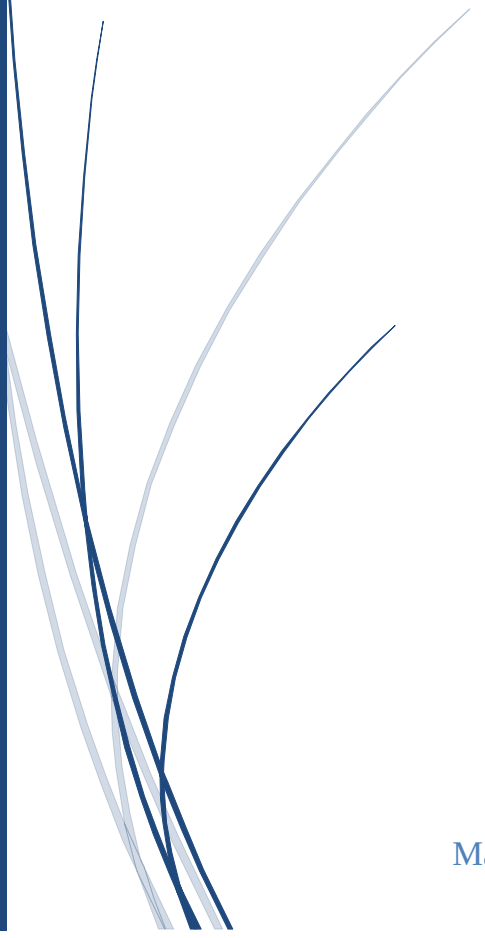
- INEXTVIR-LECTURES AND WORKSHOPS (SCHOOL 3) (07 avril 2022 - 8 juin 2022) INIA-BIOVEGEN-VATC-CSIC-DELPHY, 10 heures enregistrées par : Sciences de la Vie et de la Santé.
- INEXTVIR-ONLINE LECTURES AND WORKSHOPS (ADVANCED WORKSHOP 2) (18 octobre 2021) University of Bordeaux, 16 heures enregistrées par : Sciences de la Vie et de la Santé.
- Prepare and succeed a job interview for a recruitment In France or abroad (12 avril 2021) passage à distance ZOOM, 12 heures
- The thesis a sucessful professionnal experience (30 mars 2021) passage à distance ZOOM, 12 heures

---

270 h

Contents

GENERAL INTRODUCTION



GENERAL INTRODUCTION

Prunus species and cultivation

The genus *Prunus*, part of the family *Rosaceae* subfamily *Amygdaloideae*, consists of around 400 to 430 species including the edible and ornamental crops worldwide (Das *et al.*, 2011; Kumar *et al.*, 2021; Potter *et al.*, 2007). The most economically important species in this genus, which include peach (*Prunus persica*), European plum (*Prunus domestica*), sour cherry (*Prunus cerasus*), sweet cherry (*Prunus avium*), almond (*Prunus dulcis*), apricot (*Prunus armeniaca*) or Japanese plum (*Prunus salicina*) are used in the production of fruit, nut, and lumber (Das *et al.*, 2011; Kumar *et al.*, 2021; Potter *et al.*, 2007; Rheder, 1940). The *Prunus* genus originated in Central Asia and was introduced to Eastern Asia, Europe and North America (Janick, 2005). The highest production of stone fruits concerns peach/nectarine which is the second most economically important fruit crop in Europe. The Table 1 shows the origin and world production of the most important stone fruits (FAOSTAT 2020). China is the largest producer of peach and plum and contributes to more than 50% of the annual world production. Almonds are mainly produced by the USA, with 57% of the global annual production. Turkey is the largest producer of sweet cherry and apricot, representing almost one quarter of the global production. Sour cherry production is mainly ensured by three countries, Russia, Turkey and Ukraine, with 17%, 12%, and 11 % of the world production, respectively.

Crop	World production (million ton)	Surface (million ha)	largest producer	Production by country (Mt)	Percentage of global production by country
Peach	24.569	1.491	China	15	61.05
			Spain	1.306	5.32
			Italy	1.015	4.13
Plum	12.225	2.637	China	6.465	52.88
			Romania	0.757	6.19
			Serbia	0.582	4.76
Sweet cherry	2.609	0.445	Turkey	0.724	27.75
			USA	0.294	11.27
			Chile	0.255	9.77
Sour cherry	1.479	0.217	Russia	0.254	17.17
			Turkey	0.189	12.78
			Ukraine	0.174	11.76
Apricot	3.719	0.562	Turkey	0.833	22.40
			Uzbekistan	0.529	14.22
			Iran	0.334	8.98
Almond	4.14	2.162	USA	2.37	57.25
			Spain	0.416	10.05
			Australia	0.221	5.34

Table 1. Origin and global annual production of cultivated *Prunus* species.

Prunus virus disease and management strategies

Prunus spp. are affected by a large number of graft-transmissible pathogens (viruses, viroids, and phytoplasmas); these pathogens are responsible for more significant economic losses than generally recognized since measuring the damages caused is intricate (Hadidi *et al.,* 2011). For instance, plum pox virus (PPV) is the most destructive viral pathogen of stone fruit trees. It is the causal agent of the Sharka disease, which affects different *Prunus* species (*P. persica, P. domestica, P. dulcis, P. salicina, P. avium,* and *P. cerasus*), reducing fruit yield and quality (Cambra *et al.,* 2006; Garcia *et al.,* 2014; Scholthof *et al.,* 2011). Consequently, PPV causes considerable damage to fruit production (up to 100% premature fruit drop in the most susceptible varieties) and has a significant impact on the corresponding industries. The virus belongs to the *Potyvirus* genus in the *Potyviridae* family. Its global spread is due to the movement of infected propagation materials such as rootstock and budwood, which represents the main mean of dispersion over long distances. Moreover, it can be transmitted by at least 20 aphid species in a non-persistent manner (Gildow *et al.,* 2004; Harris, 1977). Other viruses with major impact and distributed worldwide are members of the *Ilarvirus* genus in the *Bromoviridae* family, including prune dwarf virus (PDV) and prunus necrotic ringspot virus (PNRSV). Unlike PNRSV which also infects roses and hops, PDV naturally infects only *Prunus* spp. These viruses are transmissible through seed and pollen and give rise to significant crop losses (Pallas *et al.,* 2012; Uyemoto and Scott, 1992). Apple chlorotic leaf spot virus (ACLSV) which initially was identified in apple, belongs to the genus *Trichovirus* in the *Betaflexiviridae* family and causes one of the economically important viral disease in pome and stone fruits (Myrta *et al.,* 2011). Depending on the isolate and cultivar, it can cause severe symptoms of leaf and fruit deformation, necrosis, chlorotic leaf, bark split and "pseudopox" symptom (Desvignes and Boyé, 1988) leading to yield reduction (Desvignes, 1999; Dunez *et al.,* 1972; Jelkman and Kunze, 1994; Lebas *et al.,* 2004). Nevertheless, it has a worldwide distribution because the virus is latent in some situations and can be easily disseminate by infected propagation materials (Németh, 1986).

Over the past two decades, emerging diseases caused by viruses have raised a major concern. The key determinants of virus emergence include the following: i) rapid climate changes, ii) increased global trade of agricultural commodities and planting materials, iii) low genetic diversity of the plants due to the mono-crop agricultural systems, iv) transmission of viruses from uncultivated host to crops, v) rapid evolution of viruses to

yield new variants leading to their adaptation to new hosts and vectors (Anderson *et al.*, 2004; Elena, 2014; Jones, 2009; Lefeuvre *et al.*, 2019). Viral disease management relies on two main strategies: i) cultivation of resistant varieties obtained by long-term plant breeding efforts (natural host resistance against virus or vector, genetically modified host resistance), cross-protection, and ii) prophylactic measures to restrain virus dispersion by controlling the natural vectors and using cultural methods in small scale (isolation, planting upwind, mixture of non-host plant as a barrier, rotation and others) and quarantine (control of borders), inspection, and certification (virus-free seeds or propagative material) in large scale of national and international level (Jones and Naidu, 2019; Rubio *et al.*, 2020). There is no remedy that targets viral infections, once the plant becomes infected with a virus. Therefore, one of the most efficient agronomical practices in dealing with post-viral infections is the eradication of infected plants to prevent further viral dissemination (Oberemok *et al.*, 2021). Accordingly, with regard to the globalization of trade, the legislative, biosafety, and phytosanitary controls, in particular in the quarantine and post-quarantine programs of plant material trade, are aimed at minimizing virus outbreaks instead of implementing costly and inefficient control actions after the establishment of viruses in a new environment (Jones, 2009; Jones and Naidu, 2019; Rodoni, 2009). As a consequence, the ability to identify and detect viruses remains essential for a successful management program (Maree *et al.*, 2018; Massart *et al.*, 2017; Rubio *et al.*, 2020). On the one hand, the conventional detection methods such as ELISA, molecular hybridization and PCR-based assays are only applicable when targeting well known viruses for which appropriately produced antibodies or sequence data allowing the design of probes or primers are available, permitting the development of such diagnostic tests (Clark and Adams, 1977; Ohan and Heikkila, 1993). On the other hand, such assays are unable to detect novel viruses, and the emergence of genetically divergent variants may result in the failure of accurate detection (for examples, see Gao *et al.*, 2017; Marais *et al.*, 2014).

Recent advances in High-Throughput Sequencing (HTS) provide a unique opportunity for unbiased discovery of new viruses within the need for any prior information (Massart *et al.*, 2014) and foster the efforts to characterize the virome in order to provide knowledge on the existence of novel or divergent variants and allow to identify new hosts of known viruses. Unlike eukaryotic and prokaryotic organisms, which have conserved genomic regions, such as the 18S and 16S rRNA genes that can be used in metagenomics analyses, viruses do not have such generally conserved genes or sequences (Leff *et al.*, 2017;

Roossinck *et al.,* 2015). As a consequence of such practical or technical difficulties, the current availability of phytoviral genomic sequences compared with those of plants is remarkably low (Gaafar and Ziebell, 2020; Massart *et al.,* 2019). Various methods have been used to resolve this problem by enriching the viral content in sequencing materials (Gaafar and Ziebell, 2020; Roossinck *et al.,* 2015).

Generally, depending on the different types of nucleic acids that may be targeted by various HTS approaches, they can be classified in two categories. i) the straightforward and generic sequencing approaches such as whole transcriptome analysis with total RNA sequencing (RNA-seq) that detects various forms of RNA (coding and non-coding), and the sequencing of poly-adenylated RNAs or ribodepleted total RNAs which both allow an enrichment in the coding portion of the transcriptome, ii) the approaches providing viral sequence enrichment by targeting double-stranded RNAs (dsRNAs) (Marais *et al.,* 2018a), virion-associated nucleic acids (VANA) (François *et al.,* 2018), or virus-derived small interfering RNAs (vsiRNAs) produced by plants in response to viral infection (Baulcombe, 2005; Kreuze *et al.,* 2009). These methods are further discussed below and summarized in Figure 1.

Different types of nucleic acids that can be targeted by various HTS approaches used for virus detection and discovery

Total RNA (totRNA): This is a simple, generic and non-targeted sequencing approach that is used for identifying all types of RNA and DNA viruses as well as viroids**.** This approach also potentially enables the simultaneous discovery of other, non-viral pathogens, such as phytoplasmas, and the accurate measurement of genes and transcripts abundance in transcriptomic analyses. However, high host RNA abundance may limit the sensitivity of detection of viral sequences. As a result, a high sequencing depth is required in order to be able to discover low-titer viruses, such as fruit tree viruses, which increases the cost of sequencing (Adams *et al.,* 2009; Al Rwahnih *et al.,* 2009; Roossinck *et al.,* 2015). In addition, endogenous viral elements which are derived from the ancient viral DNA integrated into the host genome are a concern, since at least some of these are known to be expressed as mRNAs and will therefore be identified. Due to the lack of annotation of such integrated viral sequence in databases, there is a risk of confusion between integrated sequences and episomal, non-integrated viral infections, reducing the sensitivity of detection for some DNA viruses with integrated forms (Aiewsakun and Katzourakis,

2015; Liu *et al.,* 2010; Massart *et al.,* 2017). On the other hand, the ability to assemble long contigs and a higher viral genome coverage (in bioinformatics analyses) compared to vsiRNA sequencing (see below) are among the main advantages of this approach (Pecman *et al.,* 2017).

Ribosomal RNA-depleted total RNA (rRNA-depleted totRNA): It is a method for the enrichment of non-rRNA transcripts. Indeed, in total RNA seq approach, as described above, the proportion of reads corresponding to ribosomal RNAs (rRNAs) can reach up to 95% of the total sequencing reads, requiring an increase of the sequencing depth. A number of methods have been developed to eliminate host rRNAs, improving detection of low-abundance transcripts and enhancing depth of coverage (Adams *et al.,* 2009; Roossinck *et al.,* 2015). This is the most popular RNA sequencing method as it allows detection of all types of viruses and viroids, including viruses without a poly-A tail.

Polyadenylated RNA (poly(A) RNA): This is an alternative method of non-rRNA enrichment. Polyadenylated RNAs, the protein-coding RNAs, are captured during the poly-A selection, therefore, rRNAs are removed, together with other non-coding RNAs. It should be noted that this approach reduces however the sensitivity of detection of viruses with non-polyadenylated genomes, since similar to cellular non-coding RNAs, the viral genome will be counter-selected during poly-A selection (Pecman *et al.,* 2017; Wu *et al.,* 2015).

Virus-derived small interfering RNA (vsiRNA): This approach is based on the small 21-24 nt RNA molecules that are produced by the plant RNA silencing system through the activity of the Dicer-like enzyme machinery (DCLs) (Hamilton and Baulcombe, 1999; Molnár *et al.,* 2005; Zhang *et al.,* 2015). siRNA can be generated in an antiviral defence against the various types of DNA or RNA viral genomes, satellites and viroids. In spite of this advantage, which provides an ability to detect different kinds of viruses including RNA and DNA viruses and viroids (Barath *et al.,* 2022; Kreuze *et al.,* 2009; Vivek *et al.,* 2020), the identification of low-titer viruses is challenging due to the complicated assembly of short length reads and therefore requires higher sequencing depth. For the same reason, full genome reconstruction and detection of recombinant events are not achievable (Kutnjak *et al.,* 2015). Similar to totRNAseq, the discrimination of the episomal viral forms from integrated ones may be problematic and require more in-depth investigations (Massart *et al.,* 2019; Pooggin, 2018). Moreover, this method is not suitable for the analysis of pooled samples often used in metagenomic studies, because the assembly

of short reads may generate chimeric sequences in case of mixed infections or in case of simultaneous presence of several variants of the same virus (Rwahnih *et al.*, 2013; Seguin *et al.*, 2014b).

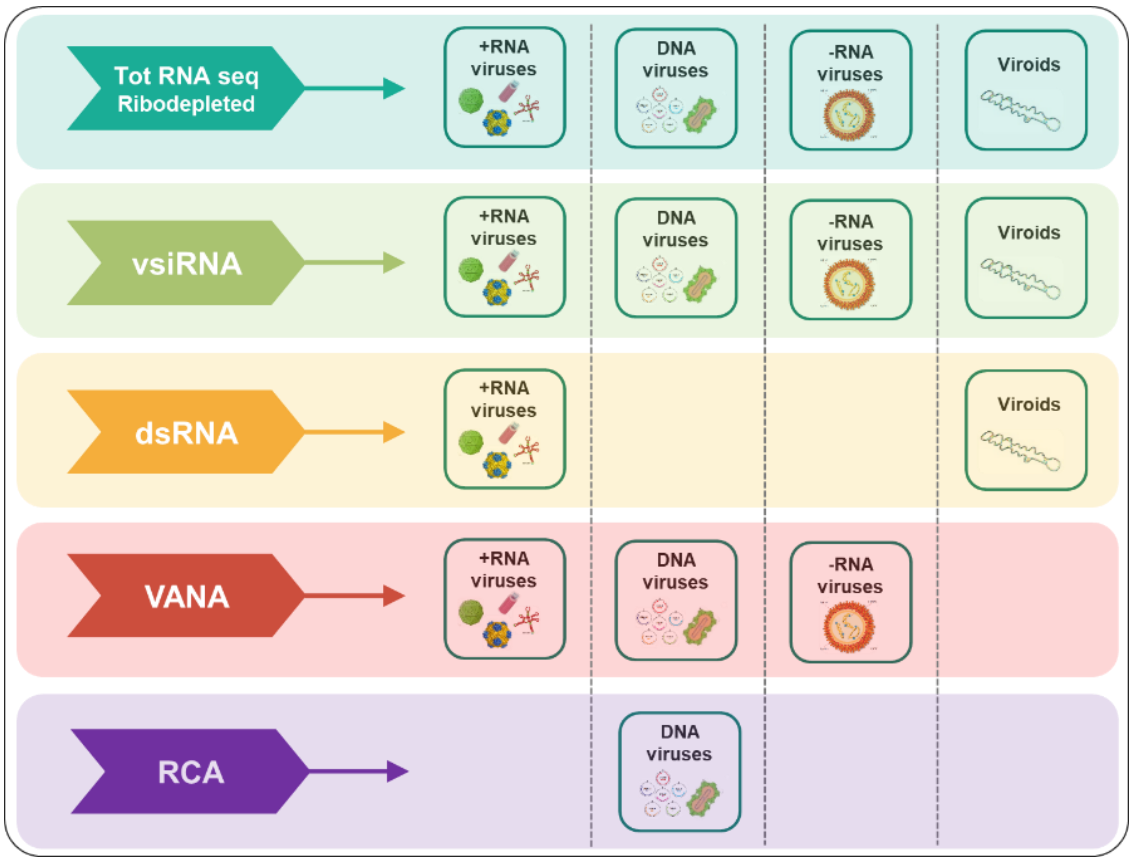


Fig 1. Different types of viral agents which can be detected through the HTS analysis of various target nucleic acids populations.

Double stranded RNA (dsRNA): the dsRNA enrichment procedure is based on the fact a large proportion of plant viruses have an RNA genome and produce dsRNA molecules during their replication, while uninfected plants do not normally produce dsRNA (Dodds et al., 1984). Therefore, accumulation of dsRNA is a hallmark of viral infection and, as such, allows targeting plant viruses including low-titer ones, as fruit tree viruses. However, it does not show the same efficiency for negative-sense RNA viruses and is not expected to allow the detection of DNA viruses, since these are not known to produce dsRNA (Kesanakurti *et al.*, 2016; Marais *et al.*, 2018a), although few DNA viruses have been identified by dsRNA-based approaches (Al Rwahnih *et al.*, 2015; Maliogka *et al.*, 2018). Depending on the yield of purified dsRNA, it can be directly submitted to HTS after

a cDNA synthesis step or submitted to random PCR amplification using unique multiplex identifier (MID) adaptors tagging prior to sequencing (François *et al.,* 2018; Marais *et al.,* 2018a; Roossinck *et al.,* 2010). This tagging has the extra advantage of allowing to pool multiplexed amplicons in a single HTS library, thus significantly reducing sequencing cost. Hence, this viral enrichment method is cost effective specially when dealing with numerous samples but, similar to PCR-based assays, susceptible to contaminations during lab operations if a random amplification step is used. The random amplification step has also often the consequence of a more uneven coverage along the viral genome than totRNA-based approaches (T. Candresse, personal communication).

Virion-associated nucleic acids (VANA) from (semi) purified viral particles: This approach is based on the semi-purification of viral particles, allowing the enrichment of protected (encapsidated) viral nucleic acids, which are subsequently submitted to a whole genome random amplification (Candresse *et al.,* 2014; François *et al.,* 2018). Similar to the dsRNA sequencing strategy, the cost of sequencing can be notably reduced due to multiplexed pooled samples using primer tags. VANA is suitable to identify a diverse range of RNA and DNA viruses (Moubset *et al.,* 2022). Apart from the advantages, there are a few drawbacks; for example, the method depends on the stability of the viral particles and is not expected to allow the identification of non-encapsidated viral nucleic acids or of viroids, though members of the *Endornaviridae* family (no particles but viral enveloped vesicles) or of the *Closteroviridae* family (relatively labile, RNAse-sensitive particles) have been detected by this approach (Bernardo *et al.,* 2017; Moubset *et al.,* 2022; Roumagnac *et al.,* 2018).

Rolling Circle Amplification (RCA): This approach is suitable only for circular single-stranded DNA viruses such as members of the *Geminiviridae* and *Nanoviridae* families and for double-stranded DNA viruses like *Caulimoviridae* family members (Idris *et al.,* 2014; Jeske 2018; Ng *et al.,* 2011).

The major differences between the approaches outlined above are related to their capacity to detect various types of viral genomes, the ease of assembly of detected viral genomes, the enrichment provided and therefore the ability to detect low titer viruses and, finally, their relative cost when analyzing a large number of samples. Thus, the choice of the best approach depends on the objective(s) of the project, the number of samples to be analyzed and the laboratory facilities which may differ from one laboratory to another. In addition, the accuracy and efficiency of the HTS can be elevated by combining different

methods. For instance, in our project, the first identification was carried out by dsRNA viral enrichment methodology (high-throughput with high multiplexing and low sequencing cost permitted by enrichment), and, if the full genome of a novel virus could not be assembled given the low sequencing depth used, interesting samples were resequenced at higher depth using a ribodepleted RNAseq method to obtain the nearly complete genome of the new virus of interest.

Data mining: an alternative, low cost approach to virus discovery

HTS approaches have been widely used in plant genomic analyses even before becoming popular in virology and the interest in such efforts has only increased. As a consequence, enormous amounts of plant-derived sequences generated in different laboratories around the world and which were not intended for the discovery of viruses are currently publicly available in sequence databases. These sequence datasets constitute a very valuable resource that can be mined to identify novel viruses or new isolates of known viruses, which might be used to design novel or more inclusive diagnostic tests. To this purpose, the Serratus tool, a cloud computing infrastructure developed recently by Edgar *et al.* (2022) allows to perform ultra-high-throughput sequence alignments against a set of query sequences at the petabase scale. I performed data-mining on publicly available data of *Prunus,* and two viral sequences were revealed to be potential new viral species.

Major bioinformatics challenges in HTS-based processing data

A part from the importance of the decent selection of HTS method as mentioned above, the choice of an efficient data analyzer among the various commercially available softwares and pipelines as well as the computing infrastructure needed may have an impact on the outcome of HTS data processing.

The most popular methods for identifying viral sequences among HTS datasets are derived from homology-based tools. The identification and classification of the HTS-derived viral sequences are based on their level of identity to the nucleotide or amino acid sequences of known viruses present in databases. There are commercial softwares such as CLC Genomics Workbench, Geneious Prime and DNASTAR's Lasergene Genomics Suite, and in-house developed pipelines such as VirAnnot (Lefebvre *et al;* 2019), Virusdetect (Zheng *et al.,* 2017), Virfind (Ho and Tzanetakis, 2014), Virtool (Gaafar *et al*.,
2021), MISIS (Seguin *et al.,* 2014a), SearchSmallRNA (de Andrade and Vaslin, 2014) or VIROME(Wommack *et al*., 2012) that are using the similarity-based approaches including

Blast, HMM, Bowtie and SOAP to analyze virome data. However, the similarity-based computational algorithms are unable to discover novel viruses that do not match anything ever seen before, defining a part of the so-called "viral dark matter" (Reyes *et al.*, 2012).

Nevertheless, to overcome this major challenge, researchers are developing innovative algorithms to identify viruses and viroids in homology-independent manner, like PFOR/PFOR2 (Zhang *et al.*, 2014), K-mer based (Baizan-Edge *et al*, 2019) and Machine Learning (ML)-based [DeepVirFinder (Ren *et al*., 2017), ViraMiner (Tampuu *et al.*, 2019) and Virnet (Abdelkareem *et al*., 2018)] approaches. For instance, in the paper I co-authored (Annex 2), VirHunter which is a ML-based tool, was trained to discriminate viral contigs from the host plant in RNAseq data that remarkably reduce the downstream work of data processing and may facilitate the discovery of very divergent novel viruses (Sukhorukov *et al.*, 2022).

Prunus-infecting viruses

As indicated above, *Prunus* species are hosts to a large range of viruses. The Table 2 lists the viruses identified prior to this project in *Prunus*, some of which were revealed as a result of HTS-based approaches. Seventy known viruses and viroids were found in *Prunus*, from which 92.9% are RNA viruses, 1.4% DNA viruses and 5.7% viroids. The Figure 2 shows the time course of *Prunus* viruses discovery for every five years period, since the first *Prunus* virus was discovered in 1933 (Atanassov, 1933). A peak of virus discovery can clearly be identified since the advent of HTS-based approaches, demonstrating the tremendous impact of HTS on virus discovery.

Table 2. *Prunus*-infecting viruses and viroids list

Viral family	Genus	Species	Acronym	HTS
Tombusviridae	*Tombusvirus*	*Carnation Italian ringspot virus*	CIRV	
		Petunia asteroid mosaic virus	PeAMV	
		Tomato bushy stunt virus	TBSV	
	Luteovirus	*Nectarine stem pitting-associated virus*	NSPaV	Bag *et al.,* 2015
		Cherry-associated luteovirus	ChALV	Lenz *et al.,* 2017
		Peach-associated luteovirus	PaLV	Wu *et al.,* 2017

Table 2. *Prunus*-infecting viruses and viroids list

Tymoviridae	*Marafivirus*	*Nectarine virus M*	NeVM	Villamor *et al.*, 2016
		Peach virus D	PVD	Igori *et al.*, 2017
Secoviridae	*Nepovirus*	*Cherry leaf roll virus*	CLRV	
		Arabis mosaic virus	ArMV	
		Cherry rosette virus	ChRV	
		Myrobalan latent ringspot virus	MLRSV	
		Apricot latent ringspot virus	ALRSV	
		Peach rosette mosaic virus	PRMV	
		Raspberry ringspot virus	RpRSV	
		Tobacco ringspot virus	TRSV	
		Tomato ringspot virus	ToRSV	
		Tomato black ring virus	TBRV	
	Cheravirus	*Cherry rasp leaf virus*	CRLV	
		Stocky prune virus	StPV	
	Stralarivirus	*Strawberry latent ringspot virus*	SLRSV	
	Fabavirus	*Prunus virus F*	PrVF	Villamor *et al.*, 2017
		Cherry virus F	CVF	Koloniuk *et al.*, 2018
		Peach leaf pitting-associated virus	PLPaV	He *et al.*, 2017
Unassigned	unassigned	*Cherry virus Trakiya*	CVT	Milusheva *et al.*, 2019
Bromoviridae	*Ilarvirus*	*Prune dwarf virus*	PDV	
		Prunus necrotic ringspot virus	PNSRV	
		Apple mosaic virus	ApMV	
		American plum line pattern virus	APLPV	

Table 2. *Prunus*-infecting viruses and viroids list

		Prunus virus I	PrVI	Orfanidou *et al.*, 2021
Botourmiaviridae	*Ourmiavirus*	*Epirus cherry virus*	EpCV	
Closteroviridae	*Ampelovirus*	*Plum bark necrosis stem pitting-associated virus*	PBNSPaV	
		Little cherry virus 2	LChV2	
	Velarivirus	*Little cherry virus 1*	LChV1	
Potyviridae	*Potyvirus*	*Plum pox virus*	PPV	
Betaflexiviridae	*Capillovirus*	*Cherry virus A*	CVA	
		Mume virus A	MuVA	Marais *et al.*, 2018b
		Apple stem grooving virus	ASGV	
	Foveavirus	*Apricot latent virus*	ApLV	
		Asian prunus virus 1	APV1	
		Asian prunus virus 2	APV2	
		Asian prunus virus 3	APV3	
		Peach chlorotic mottle virus	PeCMV	
		Cherry virus B	CVB	Yaegashi *et al.*, 2020
	Robigovirus	*Cherry green ring mottle virus*	CGRMV	
		Cherry necrotic rusty mottle virus	CNRMV	
		Cherry rusty mottle-associated virus	CRMaV	
		Cherry twisted leaf-associated virus	CTLaV	
		Cherry robigovirus 5	CRV-5	Wu *et al.*, 2019
		Cherry virus Turkey	CVTR	Çaglayan *et al.*, 2019
	Trichovirus	*Apple chlorotic leaf spot virus*	ACLSV	
		Apricot pseudo chlorotic leaf spot virus	APCLSV	

Table 2. *Prunus*-infecting viruses and viroids list

		Cherry mottle leaf virus	CMLV	
		Peach mosaic virus	PcMV	
		Peach chlorotic leaf spot virus	PCLSV	Zhou *et al.,* 2018
		Cherry latent virus 1	CLV-1	Brewer *et al.,* 2020
		Peach virus M	PVM	La Torre-Almaraz *et al.,* 2019
	Prunevirus	*Apricot vein clearing-associated virus*	AVCaV	Elbeaino *et al.,* 2014
		Caucasus prunus virus	CPrV	Marais *et al.,* 2015a
	Tepovirus	*Prunus virus T*	PrVT	Marais *et al.,* 2015b
		Cherry virus T	ChVT	Marais *et al.,* 2020
	Citrivirus	*Citrus leaf blotch virus*	CLBV	
unassigned		*Prunus yellow spot-associated virus*	PYSaV	Hou *et al.,* 2019
		Peach-associated virus 2	PaV2	Zhou *et al.,* (MN905505)
		Peach virus 1	PeV1	Zhou *et al.,* 2020
Geminiviridae	*Grablovirus*	*Prunus geminivirus A*	PrGVA	
Pospiviroidae	*Hostuviroid*	*Hop stunt viroid*	HSVd	
	Apscaviroid	*Apple scar skin viroid*	ASSVd	
		Plum viroid I	PVD-I	Bester *et al.,* 2020
Avsunviroidae	*Pelamoviroid*	*Peach latent mosaic viroid*	PLMVd	

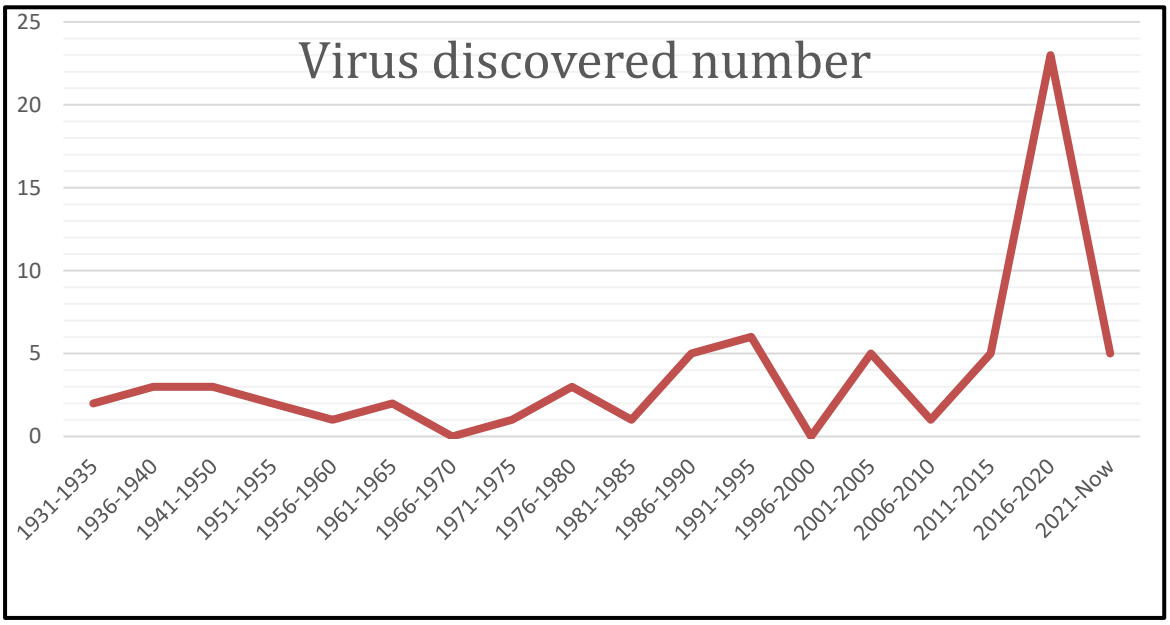


Fig 2. Number of *Prunus* viruses recorded every five years since 1930.

What happens after a new virus is discovered?

HTS was used for the first time for virus characterization and discovery in plant virology in 2009 (Adams *et al.*, 2009; Kreuze *et al.*, 2009); since then, it has largely replaced all previous sequencing technologies to identify known and unknown viruses including *Prunus* viruses (Table 1; Bag *et al.*, 2015; Brewer *et al.*, 2020; Çağlayan *et al.*, 2019; Elbeaino *et al.*, 2014 ; He *et al.*, 2017; Hou *et al.*, 2019; Igori *et al.*, 2017; Khalili *et al.*, 2020; Koloniuk *et al.*, 2018; Lenz *et al.*, 2017; Marais *et al.*, 2015a; Marais *et al.*, 2015b; Marais *et al.*, 2018b; Marais *et al.*, 2020; Milusheva *et al.*, 2019; Villamor *et al.*, 2016; Villamor *et al.*, 2017; Wu *et al.*, 2017; Wu *et al.*, 2019; Yaegashi *et al.*, 2020; Zhou *et al.*, 2018). The discovery of viruses belonging to genera previously unknown to infect *Prunus*, such as luteoviruses and fabaviruses demonstrates the efficacy of HTS in expanding our knowledge of the virome of important fruit crops. Given the accelerated use and success of HTS in virus discovery over the past decade, the molecular characterization of novel viruses has outpaced their biological characterization. This may be an emergent challenge for policy makers because it results in important data gaps when it comes to assessing the risks associated with newly discovered viruses. In the meanwhile, the authority bodies involved in the definition of quarantine and plant health certification legislation are frequently reaching cautious, constraining decisions that may not be relevant until the needed biological data is developed (Massart *et al.*, 2017).

How to proceed following the discovery of a new virus?

As described in Massart *et al.,* (2017), the identification of a new virus is only the first discovery step, which has then to be followed by biological characterization. The steps to follow after the discovery of viruses are listed as: i) confirmation of its detection and provisional taxonomic assignment, ii) sample documentation including the symptoms (if any), the origin of the sample, cultivar, time and tissue of the sampling, iii) full genome sequencing by filling the gaps or regions with low coverage and obtaining the extremities, iv) development of detection assays targeting the novel virus, v) transmission experiment and biological indexing using mechanical inoculation, grafting, vector inoculation or infectious clone preparation as well as the in-depth biological characterization. This last part in particular can take several years of gathering symptomatology, vector transmissibility, host range determination, etiology, epidemiology, vi) field survey that may be organized at small and large scales to evaluate the virus genetic diversity in order to optimize the diagnostic assays performance, in addition to generating prevalence and distribution data for the newly discovered virus. It is worth noting that in general in plant pathology, completion of Koch's postulates is a foremost biological characterization step for newly uncovered microorganisms, as it provides definite proof of the pathogenicity. However, given the very nature of viruses, plant virologists meet limitations when trying to complete Koch's postulates. This has led to efforts to propose alternative strategies for viral disease causation determination (Fox *et al.*, 2020).

Koch postulates were formulated in 1884 and refined and publish in 1890. The first postulate is the presence of the suspected microorganism in symptomatic organisms and its absence in asymptomatic ones. The second concerns the isolation of the microorganism in pure culture from symptomatic organisms. The third postulate that the cultured microorganism should cause symptoms when inoculated to a healthy organism. And finally, the fourth postulate concerns the re-isolation of the microorganism from the inoculated symptomatic organism. The main impediment to their fulfillment in virology is that viruses cannot reproduce without a host, so they cannot grow in a pure culture. Moreover, for some plant viruses, re-inoculation can also be a strong constraint, as it is the case for most of fruit tree viruses. In addition, and it is not specific for viruses, the Koch's postulates are based on "one pathogen-one disease", and don't account for asymptomatic infections and mixed infections, which are very frequent for fruit tree viruses so that Koch's postulates have been supplanted by other criteria (Bradford Hill, 1965) embedded in the

simplified hierarchical approach proposed by Fox *et al.*, (2020) for assessing causal relationship in plant virology.

Studying the *Prunus* virome

My thesis is a part of a Marie Curie European Training Network called INEXTVIR (MSCA-ITN) that has focused on identifying and characterizing plant viruses and their association to diverse habitats and plant species through HTS. INEXTVIR consists of several work packages including WP2, scanning the virome of selected agricultural crops, in which I am involved (Figure 3). The central purpose of my thesis is to enrich our knowledge of the *Prunus* virome by identifying and characterizing new *Prunus*-infecting viruses, expanding our understanding of their biology and developing diagnostic tests allowing their efficient detection. To do so, I used a dsRNA-based HTS approach to scan the virome of a large number of *Prunus* samples including 250 cultivated and 243 wild *Prunus* trees, representing a total of 26 *Prunus* species from the INRAE *Prunus* Biological Resource Center (BRC) germplasm and eight wild *Prunus* species collected in various areas. The results of this large virome characterization effort and the complete list of the *Prunus* species investigated in this study are found in **Annex 1.**

In addition to numerous known viruses, I identified novel viruses belonging to two different genera in different families. The first one is the *Luteovirus* genus in the *Tombusviridae* family, while the second is the genus *Cheravirus* in the family *Secoviridae*. The first luteovirus I identified was almond luteovirus 1, from an almond tree collected in Turkey; as no live plant material was available, biological characterization was not possible; only a molecular characterization was performed for what is the first almond-infecting luteovirus known. These results were integrated in an Annotated Sequence Record published in Archives of Virology which is presented as **Chapter I** (Khalili *et al.*, 2020).

Chapter II presents more extensive virome data and datamining efforts that led to the identification and characterization of five new *Prunus*-infecting luteoviruses in various cultivated or wild *Prunus* species such as *P. persica*, *P. armeniaca*, *P. mume*, *P. mahaleb* and *P. humilis.* I performed a molecular characterization of these viruses, developed detection assays for them, which allowed me to study the geographical distribution and the genetic diversity at the European level for three peach-infecting luteoviruses. A first biological characterization for these three luteoviruses, two previously known one and a

novel one, was performed. All these results were integrated in a paper recently accepted in Phytopathology and presented as Chapter II (Khalili *et al.*, 2022)

The work dealing with the identification and/or characterization of members of the genus *Cheravirus* is included in **Chapter III** in the form of a paper published in Viruses. I identified a novel *Cheravirus* in both wild (*P. brigantina* and *P. mahaleb*) and cultivated (*P. armeniaca*) hosts, providing valuable information on its genetic diversity and host range. In addition, I completed the genome sequence of another *Cheravirus*, stocky prune virus (StPV; Candresse *et al.*, 1998) for which only limited genomic information was available (Candresse *et al.*, 2006).

Lastly, I obtained the nearly complete genome (complete coding genome) sequence of two other known *Prunus-infecting* viruses for which no or only very partial genomic information was available, apricot latent ringspot virus (ALRSV) that was initially described in *P. armeniaca* in France (Gentit *et al.*, 2001) and myrobalan latent ringspot virus (MLRSV) described from *P. cerasifera* (Dunez *et al.*, 1971). These results are presented in **Chapter IV** as a manuscript to be submitted in Archives of virology.

Annex 2 presents the paper which I co-authored during my PhD program on developing a tool to discriminate viral sequences particularly novel RNA viruses in assemblies of sequencing datasets. VirHunter which is a trained tool, identifies the known and novel viruses in RNA sequencing data using a non-alignment-based method as indicated above (Sukhorukov *et al*., 2022).

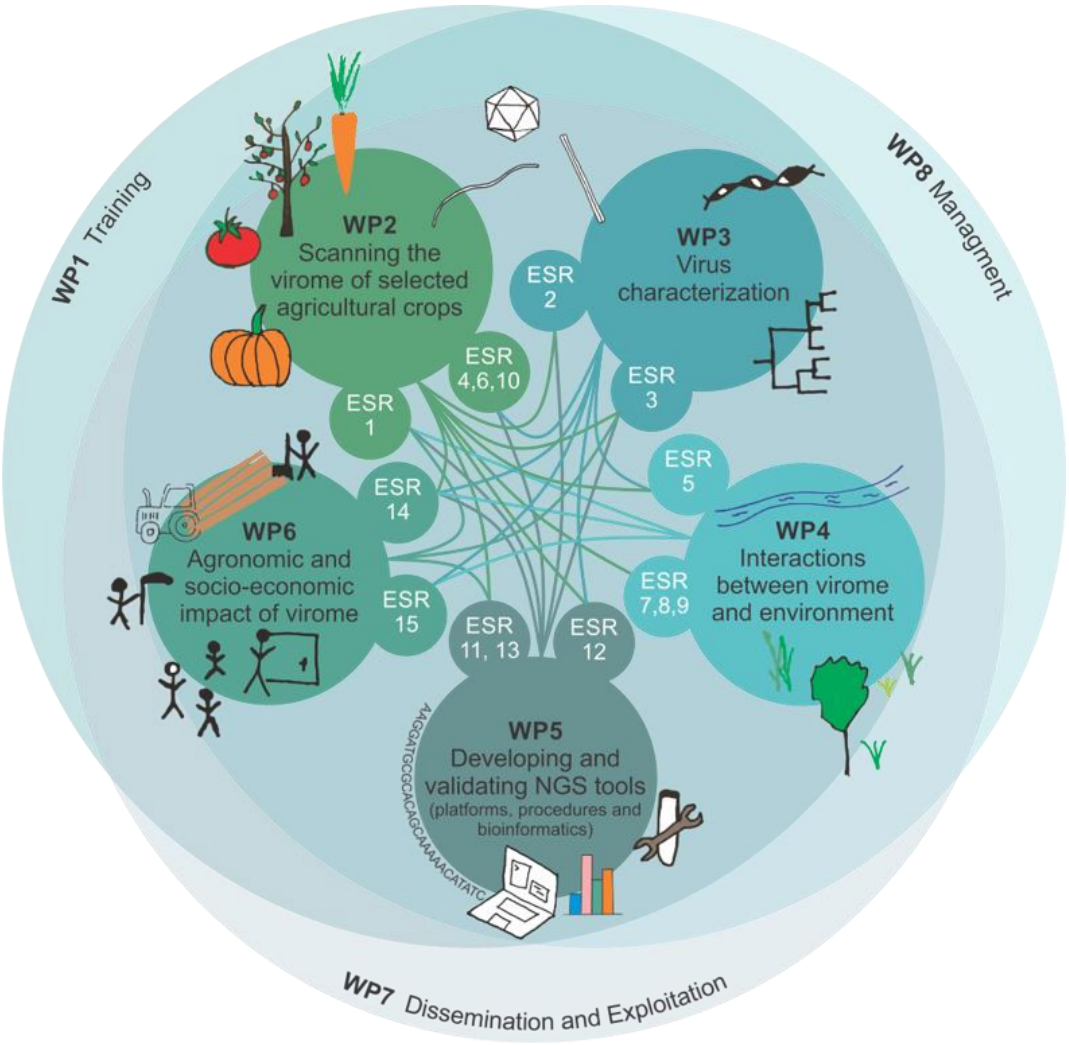


Fig 3. Overall structure of the INEXTVIR project, showing its different work packages and the links between them, together with the 15 early stage researchers (ESR) involved in the various work packages. Maryam Khalili is ESR6.

CHAPTER I

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

Archives of Virology (2020) 165:2123–2126
https://doi.org/10.1007/s00705-020-04715-2

ANNOTATED SEQUENCE RECORD

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

Maryam Khalili[1] · Thierry Candresse[1] · Chantal Faure[1] · Armelle Marais[1]

Received: 22 April 2020 / Accepted: 23 May 2020 / Published online: 2 July 2020
© Springer-Verlag GmbH Austria, part of Springer Nature 2020

Abstract

In this study, we report the complete genome sequence of a novel luteovirus detected in almond 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 replication-related proteins, the coat protein (CP), and a CP readthrough protein involved in the aphid transmission of luteoviruses. Phylogenic and pairwise distance analyses showed that this virus shares 79% and 57.8% amino acid identity 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 "*Almond associated luteovirus 1*" is proposed. To our knowledge, this is the first report of an almond-infecting luteovirus.

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 in almond.

Luteoviruses belong to the family *Luteoviridae*, which also includes 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]. NSPaV is the only luteovirus described so far with only four open reading frames (ORFs) [3]; the others contain six or seven [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 resulting cDNAs were sequenced (Hiseq3000 Illumina platform, 2x150 bp). After cleaning of reads, assembly into contigs, and BLAST-based annotation [7], a single 5-kb contig showing significant nucleotide (nt) sequence similarity to luteoviruses was identified. This contig integrated 35,411 reads (34.3% of total reads) for an average coverage of 782x and showed 71% nt sequence identity to the NSPaV genome (NC027211), which spanned from nt 39 to 4,991. The genome sequence was completed by determining the genome ends using a SMARTer RACE Kit (Takara) and has been deposited in the GenBank database under accession number MT362517.

The complete genome of the almond luteovirus is 5,047 nt long and contains four ORFs, which are homologous to the corresponding ORFs in NSPaV (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

Handling Editor: Robert H. A. Coutts.

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

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00705-020-04715-2) contains supplementary material, which is available to authorized users.

✉ Armelle Marais
armelle.marais-colombel@inrae.fr

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

(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 no ORF3a [8], ORF4 (putative movement protein), ORF6 could be identified by scanning the genome and analyzing the predicted translation products in all reading frames.

Phylogenetic analysis using complete luteovirus genomes (GenBank) confirmed that the almond virus belongs to the genus *Luteovirus* and is most closely related to NSPaV, with which it shares 71.5% nt sequence identity over the whole genome (Supplementary Fig. 1). Phylogenetic trees constructed using amino acid sequences of the RdRp or CP showed that NSPaV and the almond virus form a specific branch, with 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 sequence identity with those of luteoviruses, the highest being with NSPaV. For the CP-P5 readthrough protein, the level of sequence identity was even lower (27.2%-57.8%). Given the currently accepted molecular species demarcation criteria for luteoviruses (more than 10% aa sequence divergence in any gene product [2]), the almond virus, for which the name almond luteovirus 1 (AlLV1) is proposed, defines a novel species in the genus *Luteovirus*. In order to identify possible recombination events with related viruses, a recombination analysis was performed using RDP 4 on a multiple sequence alignment of representative isolates of recognized luteovirus species, but this failed to identify statistically supported recombination events involving AlLV1.

Fig. 2 Neighbor-joining phylogenetic trees constructed using 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 a black dot. Accession numbers are shown. The scale bars represent 10% (**A**) and 5% (**B**) amino acid sequence divergence

To estimate the incidence of AILV1 in *Prunus*, detection primers were designed (NLuteo-PolF1, 5'-GCACAT GTTTCGAGGAGATG-3'; NLuteo-PolR1, 5'- CCGTCC TCAGATCTTTCCAT-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 subjected to a two-step RT-PCR procedure [10]. Except from the sample that was already known to be positive, none of the tested samples were found to be infected by AlLV1. To the best of our knowledge, AILV1 is the first described luteovirus infecting almond. Further studies are required to estimate its distribution in stone fruits, as well as its potential pathogenicity. As for NSPaV, the absence of ORFs 3a, 4 and 6 and the presence of a shorter readthrough domain might suggest that its biology could significantly differ from that of other luteoviruses.

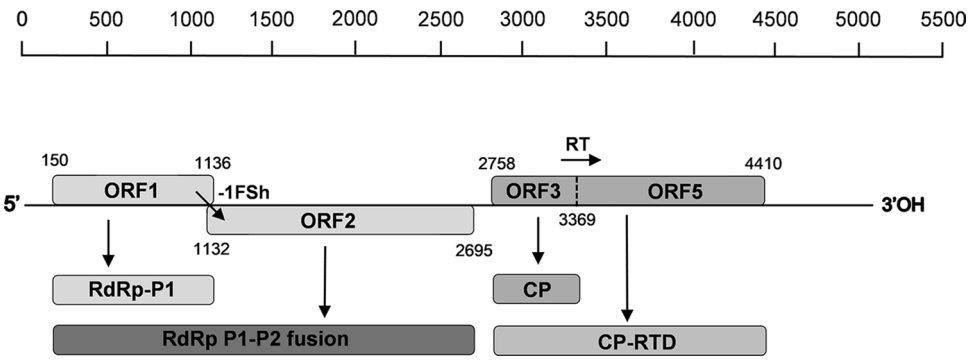
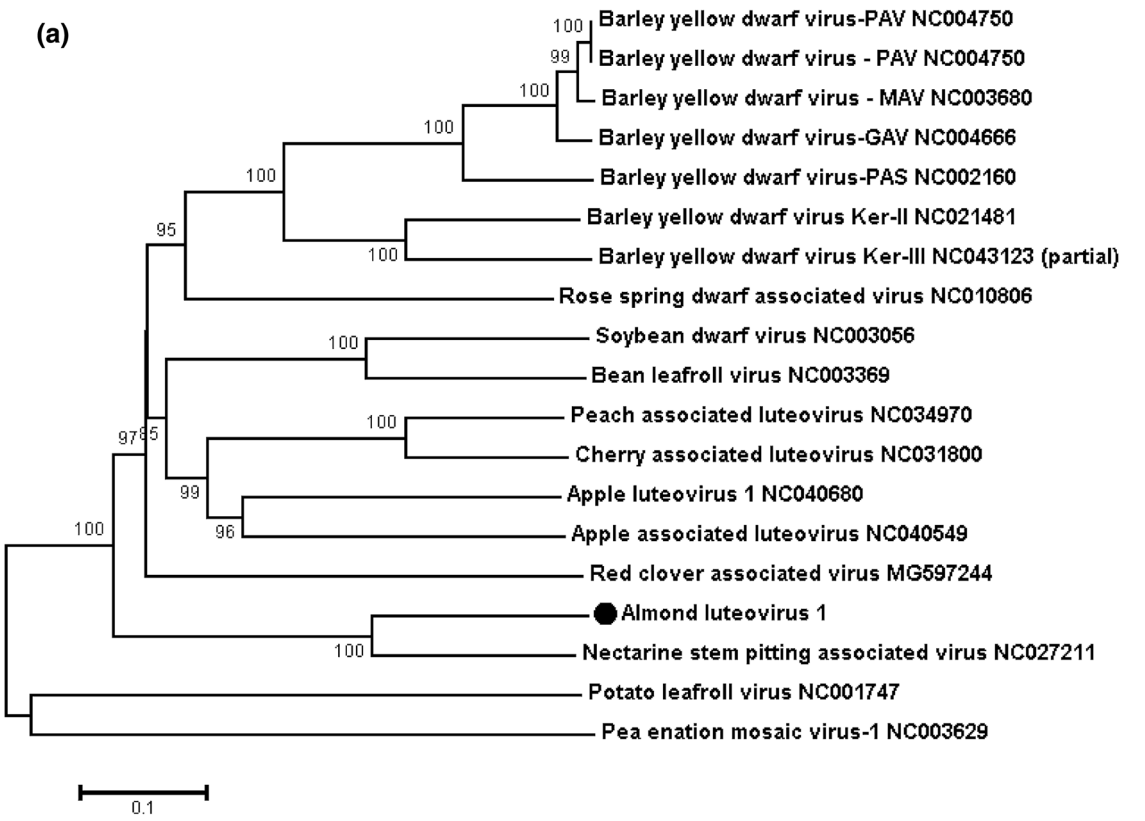
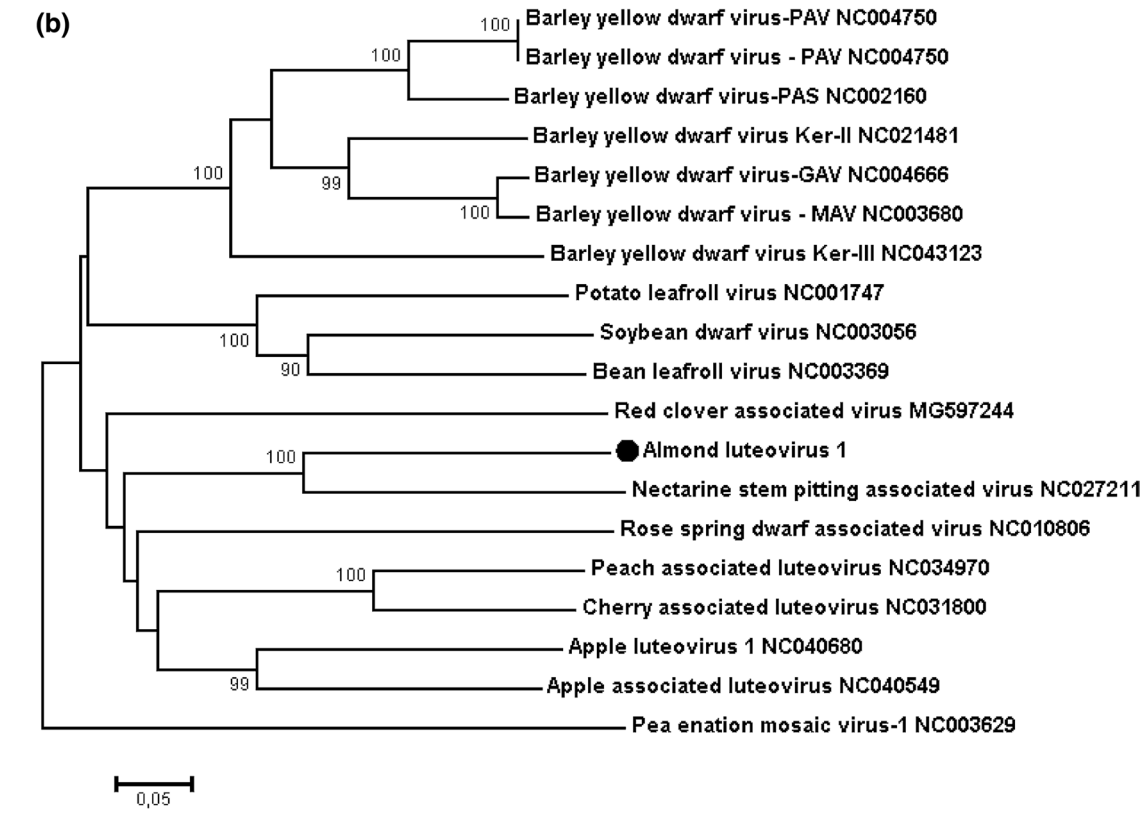


Fig. 1 Schematic representation of the genome organization of almond luteovirus 1. Boxes represent open reading frames (ORFs), and numbers indicate the coordinates of each ORF. A dashed line shows the position of the stop codon between the regions encoding the coat protein (CP) and the readthrough protein. RT, readthrough; RTD, readthrough domain; -1FSh, -1 ribosomal frameshift; RdRp, RNA-dependent RNA polymerase


(a)
Barley yellow dwarf virus-PAV NC004750
Barley yellow dwarf virus - PAV NC004750
Barley yellow dwarf virus - MAV NC003680
Barley yellow dwarf virus-GAV NC004666
Barley yellow dwarf virus-PAS NC002160
Barley yellow dwarf virus Ker-II NC021481
Barley yellow dwarf virus Ker-III NC043123 (partial)
Rose spring dwarf associated virus NC010806
Soybean dwarf virus NC003056
Bean leafroll virus NC003369
Peach associated luteovirus NC034970
Cherry associated luteovirus NC031800
Apple luteovirus 1 NC040680
Apple associated luteovirus NC040549
Red clover associated virus MG597244
Almond luteovirus 1
Nectarine stem pitting associated virus NC027211
Potato leafroll virus NC001747
Pea enation mosaic virus-1 NC003629
0.1


(b)
Barley yellow dwarf virus-PAV NC004750
Barley yellow dwarf virus - PAV NC004750
Barley yellow dwarf virus-PAS NC002160
Barley yellow dwarf virus Ker-II NC021481
Barley yellow dwarf virus-GAV NC004666
Barley yellow dwarf virus - MAV NC003680
Barley yellow dwarf virus Ker-III NC043123
Potato leafroll virus NC001747
Soybean dwarf virus NC003056
Bean leafroll virus NC003369
Red clover associated virus MG597244
Almond luteovirus 1
Nectarine stem pitting associated virus NC027211
Rose spring dwarf associated virus NC010806
Peach associated luteovirus NC034970
Cherry associated luteovirus NC031800
Apple luteovirus 1 NC040680
Apple associated luteovirus NC040549
Pea enation mosaic virus-1 NC003629
0,05

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, St. Paul, pp 271–276
2. Domier LL (2012) Family *Luteovirideae*. In: King AMQ, Adams MJ, Carstens EB, Lefkowits EJ (eds) Virus taxonomy: ninth report of the international committee on taxonomy of viruses. Elsevier Academic Press, Amsterdam, 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
8. Smirnova E, Firth AE, Miller WA et al (2015) Discovery of a small non-AUG-initiated ORF in poleroviruses and luteoviruses that is required for long-distance movement. PLoS Pathog 11(5):e1004868
9. Foissac X, Svanella-Dumas L, Gentit P et al (2005) Polyvalent degenerate oligonucleotides reverse transcription-polymerase chain reaction: a polyvalent detection and characterization tool for trichoviruses, capilloviruses, and foveaviruses. Phytopathology 95:617–625
10. Marais A, Svanella-Dumas L, Barone M et al (2012) Development of a polyvalent RT-PCR detection assay covering the genetic diversity of *Cherry capillovirus* A. Plant Pathol 61:195–204

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

CHAPTER II

The expanding menagerie of *Prunus*-infecting luteoviruses

The expanding menagerie of *Prunus*-infecting luteoviruses

Maryam Khalili[1], Thierry Candresse[1], Igor Koloniuk[2], Dana Safarova[3], Yoann Brans[4], Chantal Faure[1], Marine Delmas[5], Sébastien Massart[6], Miguel A. Aranda[7], Kadriye Caglayan[8], Veronique Decroocq[1], Pavlina Drogoudi[9], Miroslav Glasa[10,11], George Pantelidis[9], Milan Navratil[3], François Latour[4], Josef Spak[2], Jaroslava Pribylova[2], Daniel Mihalik[11], Francesco Palmisano[12], Antonella Saponari[12], Tomas Necas[13], Jiri Sedlak[14] and Armelle Marais[1] *

[1] Université de Bordeaux, INRAE, UMR BFP, Villenave d'Ornon, France

[2] Department of Plant Virology, Institute of Plant Molecular Biology, Biology Centre, Czech Academy of Sciences, České Budějovice, Czech Republic.

[3] Department of Cell Biology and Genetics, Faculty of Science, Palacký University, Olomouc, Czech Republic

[4] Laboratoire de Virologie et de Biologie moléculaire, CTIFL, Prigonrieux, France

[5] INRAE, Unité Expérimentale Arboricole, Toulenne, France

[6] Laboratory of plant pathology, TERRA, Gembloux Agro-Bio Tech, Liège University, Gembloux, Belgium

[7] Department of Stress Biology and Plant Pathology, Centro de Edafología y Biología Aplicada del Segura, CSIC, Murcia, Spain

[8] Department of Plant Protection, Hatay Mustafa Kemal University, Antakya, Hatay, Turkey

[9] Department of Deciduous Fruit Trees, Institute of Plant Breeding and Genetic Resources, ELGO-DIMITRA, Naoussa, Greece

[10] Biomedical Research Center of the Slovak Academy of Sciences, Institute of Virology, Bratislava, Slovakia

[11] Faculty of Natural Sciences, University of Ss. Cyril and Methodius, Trnava, Slovakia

[12] Centro di Ricerca, Sperimentazione e Formazione in Agricoltura "Basile Caramia", Locorotondo, Italy

[13] Department of Fruit Science, Faculty of Horticulture, Mendel University, Lednice, Czech Republic

[14] Vyzkumny A Slechtitelsky Ustav Ovocnarsky, Holovousy, Czech Republic

* Corresponding author : Armelle Marais ; E-mail : armelle.marais-colombel@inrae.fr

ORCID 0000-0003-2482-1543

Abstract

Members of the genus *Luteovirus* are responsible for economically destructive plant diseases worldwide. Over the past few years, three luteoviruses infecting *Prunus* trees have been characterized. However, the biological properties, prevalence, and genetic diversity of those viruses have not yet been studied. High throughput sequencing of samples of various wild, cultivated, and ornamental *Prunus* species enabled the identification of four novel species in the genus *Luteovirus* for which we obtained complete or nearly complete genomes. Besides, we identified another new putative species recovered from Sequence Read Archive data. Furthermore, we conducted a survey on peach-infecting luteoviruses in eight European countries. Analyses of 350 leaf samples collected from germplasm, production orchards, and private gardens showed that peach-associated luteovirus (PaLV), nectarine stem pitting-associated virus (NSPaV), and a novel luteovirus, peach-associated luteovirus 2 (PaLV2), are present in all countries, while the most prevalent virus was NSPaV, followed by PaLV. An analysis of the genetic diversity of these viruses was also conducted. Moreover, the biological indexing on GF305 peach indicator plants demonstrated that PaLV and PaLV2, like NSPaV, are transmitted by graft at relatively low rates. No clear viral symptoms have been observed either in graft-inoculated GF305 indicators, or in different peach tree varieties observed in an orchard. The data generated during this study provide a broader overview of the genetic diversity, geographical distribution and prevalence of peach-infecting luteoviruses, and suggest these viruses are likely asymptomatic in peach under most circumstances.

Keywords: HTS, Stone fruit, *Luteovirus*, geographical distribution, biological indexing

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). The ChLVA research part was financed by Academy of Sciences of the Czech Republic support (RVO60077344). The plant indexing biological tests conducted by CTIFL were funded by INTERFEL (Fresh fruit and vegetable interprofessional association). D. Safarova and M. Navratil have received support from the Ministry of Agriculture of the Czech Republic, National Agency for Agricultural Research (Project no. QK1920124). M. Glasa and D. Mihálik have received support from the Slovak Research & Development Agency (project APVV-18-0005).

Introduction

Almond and other stone fruits, such as plum, peach, sweet and sour cherry, and apricot belong to the genus *Prunus* in the family *Rosaceae*. Numerous graft-transmissible pathogens including viruses, viroids, and phytoplasmas have been described in *Prunus* and are responsible for economically important diseases, affecting the fruit industry worldwide (Hadidi and Barba, 2011). *Prunus* species host over 60 different viral and viroid species from diverse families including *Betaflexiviridae*, *Bromoviridae*, *Secoviridae*, *Botourmiaviridae*, *Closteroviridae*, *Tymoviridae*, *Potyviridae*, *Tombusviridae*, *Pospiviroidae* and *Avsunviroidae* (Hou et al. 2020; Maliogka et al. 2018; Rubio et al. 2017; Umer et al. 2019)

Members of the genus *Luteovirus* are responsible for some of the most economically important viral diseases in cereals (Miller and Rasochová 1997; Walls et al. 2019), and have also been detected in many other crops or ornamental plants including fruit trees (Bag et al. 2015; Igori et al. 2017b; Khalili et al. 2020; Lenz et al. 2017; Liu et al. 2018; Shen et al. 2018; Wu et al. 2017). The genus *Luteovirus*, formerly belonging to the family *Luteoviridae*, has recently been re-assigned to the family *Tombusviridae* (Miller and Lozier 2022). Its members have a single-stranded, messenger-sense RNA genome predicted to encode four to six (potentially eight) proteins, depending on the viral species considered (Bag et al. 2015; Hillman and Esteban, 2011; Lenz et al. 2017; Smirnova et al. 2015). Open reading frame 1 (ORF1) encodes a replication-association protein (P1), while ORF2 encodes the viral RNA-dependent RNA polymerase (RdRp). Following a -1 frameshift, RdRp is expressed as a P1-P2 fusion protein. ORF 3a, 3, 4, and 5 are translated from sub-genomic RNA1 (sgRNA1) (Domier and D'Arcy 2008; Smirnova et al. 2015). ORF3

codes for the coat protein (CP), while ORF5 is expressed as a fusion to the CP following the suppression of the leaky stop codon terminating ORF3. The small ORF3a, which is located upstream of ORF3, is translated from a non-AUG start codon (Smirnova et al. 2015) and its P3a product has been shown to be implicated in viral movement. The ORF4, which completely overlaps with the CP gene, encodes the movement protein (MP), and is translated via leaky scanning of the ORF3 start codon due to its poor context for initiation (Dinesh-Kumar and Miller 1993; Domier and D'Arcy 2008). A second subgenomic RNA, sgRNA2, likely expresses the P6 protein (Kelly et al. 1994). ORF7 encodes the putative P7 protein of unknown function and has been recently described in the genome of cherry-associated luteovirus (ChALV) (Lenz et al. 2017).

Prior to the present study, three *Prunus*-infecting luteoviruses had been described: nectarine stem pitting-associated virus (NSPaV) is the first luteovirus identified in peach (*Prunus persica*) by Bag et al in the USA in 2015 (Bag et al. 2015). Since then, NSPaV was reported naturally to infect peach in China, Hungary, South Korea, Australia (Igori et al. 2017a; Jo et al. 2017; Krizbai et al. 2017; Lu et al. 2017), and in *P. mume* (Japanese apricot) in Japan (Candresse et al. 2017). Furthermore, it has been experimentally shown that NSPaV can infect *P. avium* (sweet cherry) and *P. tomentosa* (Nanking cherry) (Villamor et al. 2016). Later, ChALV was characterized in *P. avium* and *P. cerasus* from the Czech Republic (Lenz et al. 2017). Peach-associated luteovirus (PaLV) was initially described in the USA from peach material imported from Georgia and Spain (Wu et al. 2017) and has since been reported, again from peach, in China, South Korea, Italy, and Hungary (Barath et al. 2022; Igori et al. 2017b; Sorrentino et al. 2018; Zhou et al. 2018).

Luteoviruses generally have aphid vectors (Ali et al. 2014) but this has not yet been verified for *Prunus*-infecting luteoviruses.

The association between *Prunus* luteoviruses and symptoms in their hosts is still unclear. Even for NSPaV that was initially isolated from nectarine trees showing extensive pitting on their woody cylinder (Bag et al. 2015), the authors pointed out the difficulty to correlate the symptoms with the virus presence. In addition, in another study (Villamor et al. 2016), NSPaV was detected together with a marafivirus in multiple nectarine and peach trees, suggesting a complex or non-existent relationship between the stem pitting symptoms and the two viruses. The same conclusion can be drawn from two studies on the PaLV pathogenicity (Sorrentino et al. 2018; Wu et al. 2017). Similarly, in the case of ChALV, it was not possible to draw clear conclusions due to the presence of other co-infecting viruses (Lenz et al. 2017).

The discovery of stone fruit tree viruses using high throughput sequencing (HTS) approaches has sped up over the last two decades (Hou et al. 2020; Maliogka et al. 2018; Rubio et al. 2017). But one of the limitations of these studies is that there are plenty of novel viruses discovered for which no or only very limited information is available on their biological properties and prevalence to assess the potential risk they might pose to the trees (Massart et al. 2017).

Using the HTS approach, we identified four new *Prunus*-infecting luteoviruses in the present study. A fifth one was discovered following a screening approach of publicly available *Prunus* RNA-Seq Sequence Read Archive (SRA) data. All five novel *Prunus*-infecting luteovirus species were characterized at the molecular level. Besides, we evaluated the peach-infecting luteoviruses for their graft transmissibility and, as a part of

a European field survey of peach trees, their prevalence, distribution, and genetic variability.

Material and methods

Plant material origin. Fifty peach tree (*P. persica*) accessions introduced between 1937 and 2010 from different countries in the *Prunus* INRAE Biological Resource Center (BRC Toulenne, France) were indexed by HTS. For each accession, five leaves from different parts of the tree were collected in June 2019 and pooled in equal ratios, constituting the sample analyzed by HTS. In addition, a few trees belonging to various *Prunus* species were also analyzed by HTS. For these trees, leaf samples were collected over the 2013-2021 period in various countries, regardless of the presence of symptoms (Table 1). Until used, fresh leaf tissues were either desiccated over anhydrous $CaCl_2$ (Sigma Aldrich Chimie, Saint-Quentin-Fallavier, France) and stored at room temperature or at -80°C.

To evaluate the prevalence of the luteoviruses identified in *P. persica,* samples from peach trees originating from seven European countries (in addition to the 50 French samples cited above) were obtained either from germplasm collections or production orchards. Between 26 and 51 trees were thus sampled depending on the country: Belgium (26), Greece (30), Czech Republic (43), Italy (51), Slovakia, Spain, and Turkey (50 each). These 350 peach trees were analyzed individually for the presence of some of the *Prunus*-infecting luteoviruses, including NSPaV, PaLV (known luteoviruses) and PaLV2 (a new luteovirus), while MaLV and PhaLV (the novel luteoviruses characterized in this work) where analyzed as pooled samples.

Double-stranded RNA extraction, library preparation, and sequencing. Double-stranded RNAs (dsRNA) were purified from pooled leaves (S1, S3, S4, and S7 samples, Table 1) by batch chromatography on cellulose CC41 (Whatman) as described (Marais et al. 2018), and converted to cDNA using LDF primers (François et al. 2018, Supplementary Table S1) and SuperScript™II Reverse Transcriptase according to manufacturer's instructions (Invitrogen/Fisher Scientific, Illkirch, France). Each cDNA preparation was subjected to a random PCR amplification using multiplex identifier (MID) adaptors (François et al. 2018, Supplementary Table S1), allowing to sequence all the samples in a multiplexed format. Five microliters of cDNA were amplified according to Marais et al. (2018) in a 50 µl reaction containing 10× buffer, 4 mM dNTPs, 1 µM primer MID tag, 1.25 U Dream Taq DNA polymerase (Thermo Fisher Scientific). Random PCR amplification was performed for one cycle of 94°C for 1 min; 65°C for 0 s; 72°C for 45 s, and 40 cycles of 94°C for 0 s; 45°C for 0 s; 72°C for 5 min, and 1 final cycle of 5 min at 72°C and 5 min at 37°C. Following the purification of the PCR products using a MinElute PCR Purification Kit (Qiagen SAS France, Courtaboeuf, France), PCR products were pooled equimolarly before being sent for Illumina sequencing on a Hiseq3000 platform (2x150 bp) [outsourced at the GetPlage INRAE platform (Toulouse, France) or Azenta (Leipzig, Germany)].

Alternatively, dsRNAs were extracted from 1 g of leaf tissue (S5 and S6 samples, Table 1) using the CF11 cellulose protocol of De Paulo and Powell (DePaulo and Powell, 1995) and converted into double-stranded cDNA using the Maxima H Minus Double-Stranded cDNA Synthesis Kit (Thermo Scientific, Waltham, MA, USA). The sequencing library was prepared using the Illumina compatible MuSeek Library Preparation Kit (Thermo

188 Scientific) with the double-stranded cDNA as input material and sequenced using a
189 HiSeq2500 system in 1x100 bp mode (SEQme.eu, Dobříš, Czech Republic).

190 **High throughput sequencing of total RNAs.** Total RNAs were extracted from
191 desiccated leaves of the *P. mahaleb* sample (S8, Table 1) using a modified CTAB
192 procedure (Chang et al. 1993), reverse-transcribed, ribodepleted, and sequenced
193 (HiSeq3000 2x150 bp). Alternatively, total RNAs were isolated from four leaves (100 mg)
194 of the *P. armeniaca* sample (S2, Table 1) using the Plant/Fungi Total RNA purification kit
195 (Norgen Biotek). Purified RNAs were ribodepleted using the QIAseq FastSelect-rRNA
196 Plant Kit (Qiagen) and a library prepared using the NEBNext Ultra II Directional RNA
197 Library Prep Kit before being sequenced in a multiplex run (NovaSeq6000, 2x 161 bp,
198 Institute of Experimental Botany, CAS, Olomouc, Czech Republic).
199 **HTS data analyses.** Sequencing reads were quality-trimmed using CLC Genomic
200 workbench software version 21.0.3 (Qiagen) or Geneious Prime (Biomatters Ltd,
201 Auckland, New Zealand). Following *de novo* assembly of contigs, a BlastX analysis was
202 performed against the GenBank non-redundant (nr) protein database restricted to viruses,
203 to identify viral contigs. Sequence datasets were also analyzed by mapping trimmed reads
204 on a collection of reference viral genomes (min length fraction=0.9; min similarity
205 fraction=0.7). The initially identified luteoviral contigs were then scaffolded (if needed) and
206 extended by multiple rounds of mapping using residual reads in CLC Genomics
207 Workbench to generate nearly complete genomic sequences. For isolates of known
208 viruses, no further effort was made to fill small internal gaps or the genome terminal ends,
209 but for newly discovered viruses, the genomic sequences were completed as described
210 below.

Completion of the genome sequence of the identified new viruses. In order to obtain the complete genome sequence of the newly discovered viruses peach-associated luteovirus 2 (PaLV2) and mume-associated luteovirus (MaLV), Rapid Amplification of cDNA Ends (RACE) experiments were carried out for both 5' and 3' ends using the SMARTer® RACE 5'/3'Kit (Takara Bio Europe SAS, Saint-Germain-en-Laye, France) and heat-denaturated (10 min at 99°C) dsRNAs as a template, following the manufacturer's instructions. Alternatively, the cherry luteovirus A (ChLVA) genome termini amplification was done using total RNAs and 5'- and 3'-RACE kits following the manufacturer's recommendations (Invitrogen, Waltham, MA, USA) with the virus-specific primers (Supplementary Table S1). Prior to the 3'-RACE, total RNAs were polyadenylated using ATP and poly(U) polymerase following the manufacturer's recommendations (NEB, Ipswich, MA, USA). Obtained RACE products sequenced (Eurofins Genomics, Ebersberg, Germany) using the virus-specific primers. All specific RACE primers used were designed from the sequence of the identified viral contigs and are listed in Supplementary Table S1.

Data mining. To uncover potential new luteoviruses in publicly available RNA-Seq data, we performed an analysis on SRA using Serratus, an open-source cloud computing infrastructure (Edgar et al. 2022) that seeks the closest matched SRA sequences to an input virus using a 102 amino acid (aa) viral RNA-dependent RNA polymerase sequence (RdRp palmprint). The sequence of the contig thus identified from a *Prunus humilis* SRA from China (SRR12442710) has been deposited in GenBank under the BK061315 accession number.

Phylogenetic, recombination and genetic population analyses. Multiple alignments of nucleotide (nt) or amino acid (aa) sequences were performed using the ClustalW

234 program (Thompson et al. 1994) implemented in Mega 7 (Kumar et al. 2016).
235 Phylogenetic trees were constructed using the neighbor-joining technique with strict nt or
236 aa distances and randomized bootstrapping to evaluate branching validity. Mean
237 diversities, and genetic distances (p-distances calculated on nt or aa identity) were
238 calculated using Mega 7. The RDP4 program (Martin et al. 2015) was used to search for
239 potential recombination events in the luteovirus genomic sequences obtained in this
240 study.

241 **Molecular detection of luteoviruses by RT-PCR for HTS validation, prevalence**
242 **determination, and genetic diversity analysis.** Total nucleic acids (TNA) were
243 extracted from *Prunus* leaves according to the procedure 1 described in Foissac et al.
244 (2005). The virus-specific primers were designed using the identified viral contigs
245 sequences (Supplementary Table S1) and used to detect the targeted viruses by two-step
246 RT-PCR assays. Briefly, TNA were first submitted to a reverse transcription initiated by
247 pdN_6 primers and using RevertAid H Minus First Strand cDNA Synthesis Kit (Thermo
248 Scientific). Complementary DNAs were then amplified using specific primers and either
249 the Dream Taq DNA polymerase (Thermo scientific) or the Advantage 2 polymerase mix
250 (Takara Bio Europe). Amplified products were analyzed by agarose gel electrophoresis
251 and Sanger sequenced on both strands (Eurofins). The PCR product sequences have
252 been deposited in the GenBank database under the accession numbers ON637949 to
253 ON638176.

254 **Graft transmission to GF305 peach indicator seedlings.** Based on their virome
255 composition, 24 peach trees of the INRAE *Prunus* BRC were selected for biological
256 indexing. New flush twigs were collected in June 2021 and kept at 4°C prior to chip-
257 budding on GF305 peach indicator seedlings. The grafting assays were carried out using

two twigs as budwood for every peach accession and 3-10 grafted seedlings per accession depending on twig size. Each grafted seedling was grafted with two bark pieces. In total, 199 GF305 plants were graft-inoculated in addition to five negative controls self-grafted using healthy GF305 plants free of *Prunus* viruses and viroids. The grafted plants were maintained under controlled greenhouse conditions for six months to monitor the appearance of symptoms. After the first cycle of observation, the plants were stored at 2°C to induce artificial dormancy. After 3.5 months of dormancy, the graft-inoculated plants were cut back to 30 cm high and placed again in greenhouse for a second cycle of observation.

The presence of the various viruses in the grafted GF305 seedlings was assessed by testing leaves and using specific RT-PCR assays. The identity of the amplicons was confirmed by Sanger sequencing. Graft transmissibility rate was assessed by sampling individually each inoculated GF305 plant for 10 accessions, with 4-10 grafted seedlings per accession. For the other 14 accessions, grafted GF305 seedlings (3-10 grafted plants) were not tested individually but as a pool of leaves from all grafted plants for each accession. A positive reaction would indicate that at least one of the grafted trees had acquired the virus.

Results

Identification of four novel *Luteovirus* species and of new *Prunus* hosts for NSPaV.

As part of a systematic effort to explore the virome of *Prunus* species, dsRNAs or total RNAs extracted from a wide range of *Prunus* samples were analyzed by HTS. Following reads quality trimming, *de novo* assembly and contigs annotation based on BlastX

analysis, several contigs with similarities to *Luteovirus* genus members were identified in a range of samples. Contigs of interest were then assembled into scaffolds and extended by successive rounds of residual reads mapping to yield finalized contigs spanning in many cases near-complete genomes. A detailed analysis of the assembled genomes (see below) revealed that four of them shared less than 90% aa identity in at least one of their encoded proteins with known luteoviruses, which is below the molecular demarcation threshold (10% aa divergence in any gene product) for new species in the genus *Luteovirus* (Hillman and Esteban, 2011). Overall, four sequences representing potentially four new species were thus identified in samples from *P. mume* (S1), *P. persica* (S4), *P. cerasus* (S6), and *P. mahaleb* (S8) (Tables 1 and 2, Supplementary Table S2), with the proposed names of mume-associated luteovirus (MaLV), peach-associated luteovirus 2 (PaLV2), cherry luteovirus A (ChLVA), and Prunus mahaleb-associated luteovirus (PmaLV), respectively. The genomic sequences of the PaLV2, MaLV, and ChLVA isolates were completed by filling internal gaps by PCR if needed, and by determining 5' and 3' genome ends by RACE. The 5,822 nt contig for PmaLV, lacking only 10 nt and 40 nt at the 5’ and 3’ ends respectively, as judged from a comparison with the most closely related luteovirus, ChALV (NC_031800) was not completed. The corresponding genome sequences have been deposited in the GenBank database under the accession numbers ON408234 (PaLV2), ON408236 (MaLV), ON408238 (PmaLV) and ON146357 (ChLVA) (Supplementary Table S2). The number of HTS reads mapped to each genome and the average genome coverage are presented in Supplementary Table 2. In addition to these complete genomic sequences, near-complete genomes were also obtained from other *Prunus* samples, allowing the identification of divergent variants of MaLV in *P. armeniaca* (sample S2) and *P. incisa* (sample S3), of ChLVA in a second *P. cerasus* from cv Cigany

(S6) and of a variant of NSPaV from *P. cerasus* (S7) (Table 1, Supplementary Table S2). NSPaV infection was also identified in a *P. brigantina* sample (S9), but the low viral concentration precluded the assembly of large contigs. The infection status of all samples was in all cases validated using virus-specific RT-PCR assays and sequencing of the amplicons. The near-complete genomic sequences of MaLV and NSPaV isolates have been deposited in GenBank under the following accession numbers: ON408233 (NSPaV, *P. cerasus*), ON408235 (MaLV, *P. incisa*) and ON408237 (MaLV, *P. armeniaca*), (Supplementary Table S2).

Identification of a novel *Luteovirus* species from publicly available *Prunus* RNASeq data. To uncover other luteoviruses infecting *Prunus*, the Serratus tool (Edgar et al. 2022) was used with RdRp sequences of PaLV2 and MaLV, two of the four newly identified viruses in this study, as queries. At the species level, only one RNAseq SRA (*P. humilis* from China, SRR12442710) was identified, with a contig showing 83% aa identity in the highly conserved RdRp motif with both queries, indicating that this sequence likely represents a new species in the genus *Luteovirus* of this tentative agent. The SRA dataset was downloaded and, following *de novo* assembly using CLC Genomics Workbench, a large contig of 5,202 nt (nearly full-length, in comparison to other *Prunus*-infecting luteoviral genomes) was identified. This contig shows only 48-73% nt identity with any known luteovirus species, suggesting this isolate belongs to a novel species in the genus *Luteovirus*. The sequence of this contig has been deposited in GenBank (BK061315) and the name Prunus humilis-associated luteovirus (PhaLV) is proposed for the corresponding novel species (Supplementary Table S2).

Molecular characterization and phylogenetic affinities of the five novel luteoviruses. As indicated above, the full-length genomic sequences of PaLV2 (S4), MaLV (S1) and ChLVA (S6) isolates were determined and shown to be respectively 5,780 nt, 5,748 nt, and 5,726 nt. A near-complete genome of 5,822 nt is also available for PmaLV (S8), together with near-complete genomes of the MaLV isolates from *P. armeniaca* (5,733 nt, S2) and *P. incisa* (5,705 nt, S3), as well as a near-complete genome for a second ChLVA isolate from *P. cerasus* cv Cigany (5,689 nt, S6). The NSPaV scaffold detected in *P. cerasus* represents very likely the complete genome of this isolate (4,993 nt, S7). The near-complete genome assembled from SRA data for PhaLV (5,202 nt) could obviously not be completed by the RACE experiment but the available sequence covers completely the virus open reading frames (ORFs).

The genomes of ChLVA, MaLV, PaLV2, PhaLV, and PmaLV encode six to eight ORFs and have an organization similar to those of other members of the genus *Luteovirus* (Table 2 and Fig. 1A). The main variability observed concerns the short P6 and P7 ORFs, which are missing in some viruses or isolates: ORF6 is absent in one isolate of MaLV (S3 from *P. incisa*) and PaLV2 (Fig. 1A) and ORF7 is absent in most *Prunus*-infecting luteoviruses with the exception of PaLV, ChALV and ChLVA (Table 2). Surprisingly, unlike the previously reported reference NSPaV isolate from *P. persica*, the NSPaV isolate reported here from *P. cerasus* has an ORF6. There is thus both between-species and within-species presence-absence variability for these two small putative ORFs. The second main divergence from the typical genomic organization for luteoviruses concerns NSPaV, with the *P. cerasus* isolate lacking an ORF4 and an ORF3a and having a shorter ORF5, as previously reported for other NSPaV isolates and for almond luteovirus 1 (AlLV1) (Bag et al. 2015; Khalili et al. 2020). ORF3a is also missing in the genome of PmaLV (Table 2).

A phylogram constructed using a whole-genome sequence alignment of all *Prunus*-infecting luteoviruses divides them into two clades (Fig. 1B). While NSPaV and AlLV1 form a distinct clade, the rest of the *Prunus*-infecting luteoviruses groups together with a high bootstrap support. Interestingly, the ORF encoding the MP is systematically present in luteoviruses belonging to this latter group, whereas it is absent in NSPaV and AlLV1. Phylogenetic trees based on the sequences of P1-P2 and P3-P5 fusion proteins were also generated and showed the same clustering pattern (Supplementary Fig. S1).

To precisely determine the phylogenetic affinities between *Prunus*-infecting luteoviruses, pairwise comparisons for the P1-P2 and P3-P5 proteins were performed (Supplementary Fig. S2). Whatever the luteovirus species and the protein considered, the level of aa identity was less than 90%, with the exception of PmaLV and ChALV which show 95% aa identity in the P1-P2, but only 88% in the P3-P5, supporting the notion that they should belong to distinct species. In addition, viral isolates identified as belonging to the same species, *i.e* NSPaV-*P. cerasus*, MaLV-*P. incisa*, MaLV-*P. armeniaca*, and ChLVA-Cigany, displayed more than 90% of aa identity in their various proteins with those of their respective reference isolates, thus confirming their taxonomic assignation (Supplementary Fig. S2).

To determine whether recombination has played a role in the evolution of the newly identified luteoviruses, an RDP4 recombination analysis was performed on a full genome multiple alignment. No recombination signature with significant support involving *Prunus*-infecting luteoviruses was detected (data not shown).

HTS virome characterization of peach accessions in INRAE *Prunus* BRC. As part of the *Prunus* virome characterization effort, a total number of 50 *P. persica* accessions were

individually analyzed by dsRNA-based HTS indexing. Upon demultiplexing and quality trimming steps, an average of 1.5 million reads (range 0.24 to 5 million reads) were obtained per individual sample. Apart from infrequent infections involving well-known peach-infecting viruses such as apple chlorotic leafspot virus (*Betaflexiviridae*), prunus necrotic ringspot virus (*Bromoviridae*), little cherry virus 1 (*Closteroviridae*), plum bark necrosis stem pitting-associated virus (*Closteroviridae*) and peach latent mosaic viroid (*Avsunviroidae*), BlastX analysis of the assembled contigs revealed that NSPaV, PaLV, and the newly discovered PaLV2 showed high prevalence in the peach accessions analyzed. The HTS reads datasets were also analyzed by mapping trimmed reads on reference luteovirus genomes and the results were validated by RT-PCR using corresponding virus-specific detection primers. Altogether, the results showed that 96% of the 50 peach accessions are infected by NSPaV, compared to 38% for PaLV and 54% for PaLV2.

Resampling of the luteovirus-infected trees was performed in 2021, two years after the original sampling, as well as observations for any leaf or wood symptoms. No clear symptoms of viral infection could be identified in the field-grown trees and, in particular no symptoms of stem pitting on their bark or woody cylinder. RT-PCR testing of leaf samples showed that viral infection was detected again in 71%, 77%, and 87% of the trees initially found infected in 2019 by PaLV, PaLV2, and NSPaV, respectively, indicating that infection by any of the three viruses could persist over a 2-year period but also that no further spread had apparently occurred. In order to evaluate the distribution of the viruses within individual trees, individual leaves taken from five different parts of the canopy of three trees were separately tested by virus-specific RT-PCR. NSPaV, PaLV and PaLV2 were detected in 9/10, 5/5, 10/10 individual leaves, respectively.

Geographical distribution, prevalence and genetic diversity of peach infecting luteoviruses. As shown above, three luteoviruses (NSPaV, PaLV, and the new PaLV2) had high prevalence in the French peach BRC samples. To study the geographical distribution of luteoviruses in peach in Europe, 350 peach samples originating from seven countries including Belgium, Czech Republic, Greece, Italy, Slovakia, Spain, and Turkey were collected trying to maximize varietal diversity and without taking into consideration the presence of potential viral symptoms. All samples were tested by RT-PCR using virus-specific primers individually as above (Supplementary Table S1). Amplicons from positive samples were subjected to direct Sanger sequencing in order to confirm the specificity of the amplification and assess the genetic diversity of the various viruses (see below). Remarkably, all three viruses (NSPaV, PaLV and PaLV2) were identified in peach samples from all seven countries, their incidences are shown in Table 3. On the contrary, all tested peach samples were found negative for the *Prunus*-infecting luteoviruses not reported so far in peach, including MaLV, and PhaLV. The most prevalent virus is NSPaV with an average prevalence of 66% [range 27% (Italy) to 100% (Czech Republic)], followed by PaLV with an average prevalence of 40% [range 6% (Turkey) to 88% (Slovakia)] and finally PaLV2 with an average prevalence of 14% [range 3% (Greece) to 54% (France)]. In total, 216 different varieties out of 256 varieties (71 samples had no information available on their variety) were found to be infected by either NSPaV or PaLV or PaLV2. A subset of amplicons (up to 15 per virus and per country) were submitted to Sanger sequencing and the nucleotide sequences, together with all available reference sequences, were used to construct a phylogenetic tree for each virus (Fig. 2). A total of 103 amplicon sequences were thus generated for NSPaV, 87 for PaLV, and 38 for PaLV2. The overall mean nt diversities in the short PCR fragments used for detection (3.7% +/-

0.006% for NSPaV and PaLV2, 7.1% +/- 0.009% for PaLV), as well as the topology of the trees (Fig. 2) show a generally limited genetic variability between isolates originating from different countries.

Graft-transmissibility of peach-infecting luteoviruses. In order to provide some insights into the biology of the peach-infecting luteoviruses, their graft transmissibility to GF305 peach seedling indicators as well as the symptoms induced were evaluated using samples from the INRAE *Prunus* BRC for which a full HTS viral indexing had been confirmed by specific RT-PCR assays. This included accessions with single or multiple luteoviral infections, with or without mixed infections with other well-known *Prunus* viruses or viroids (see above). A 100% transmission rate was observed for other co-infecting viruses and viroids, including PNRSV, ACLSV and PLMVd, confirming the efficiency of the transmission assay (Table 4). On the other hand, for 10 accessions based on the individual testing of inoculated GF305, the overall rate of transmission of NSPaV was estimated at 55.4%, while that of PaLV was 30% and that for PaLV2 at 8.3% (Table 4). The rates of transmission from individual accessions were also quite variable but could not be easily correlated with the infection status (single or multiple infections) of the original peach accession. In GF305 grafted with the remaining 14 accessions tested as composite pools of leaves, NSPaV was detected in 5 out of 14 pools, whereas PaLV was only detected in 1 out of 4 pools and PaLV2 was not detected in the 2 relevant pools. A visual inspection of the graft-inoculated GF305 plants was performed six months after grafting. As expected, all GF305 plants grafted with the accession co-infected with ACLSV displayed the expected dark green sunken mottle symptoms typical of ACLSV in this widely used indicator (data not shown). On the contrary, most (7/9) of the GF305 plants

grafted with the other accessions showing various luteoviral combinations revealed no visible symptoms on any of the grafted plants (Table 4). For two accessions (S8278 X6Y75 O3 and S3527 X2Y16 O1, Table 4), symptoms of leaf chlorosis, reddening, or deformation could be observed in respectively 3/9 and 2/9 grafted plants. After 3.5 months of cold-induced dormancy, a second round of observation was conducted (Table 4). In the case of S8278 X6Y75 O3, no symptoms were expressed during this second growth cycle. For S3527 X2Y16 O1, leaf reddening or chlorosis were observed again in 2/9 plants but these symptoms were not correlated with NSPaV infection since positive trees were either symptomatic or asymptomatic. In addition, one case of stem necrosis (S1161 X7Y8 O3) and one case of leaf chlorosis/reddening (S5555 X4Y67 O2) were observed (Table 4), but these were not associated with NSPaV infection.

Discussion

This study describes five novel luteoviruses identified from different *Prunus* species. Compared to the previously reported three *Prunus*-infecting luteoviruses (NSPaV, PaLV and ChALV), these results provide further evidence of the power of HTS-approaches for the discovery of unknown viruses, even in situations of latent or mixed infections. However, the *in silico* discovered PhaLV should be considered with caution since it has not been possible to experimentally validate its presence in this host. However, the fact that PhaLV could also be identified in RNASeq data independently generated (PRJNA683804) is in favor of the existence of PhaLV in *P. humilis*.

Compared with all other known luteoviruses, the five viruses characterized here share less than 90% of aa sequence identity in at least one of their proteins, which is the currently accepted molecular species demarcation criteria in the genus *Luteovirus* (Hillman and Esteban, 2011). Phylogenetic analyses demonstrate their close affinities with previously described ChALV and PaLV with which they form a monophyletic clade. We also identified divergent isolates for MaLV, ChLVA, and NSPaV. The discovery of isolates of MaLV in *P. mume*, *P. incisa*, and *P. armeniaca* indicates the ability of this virus to infect a range of ornamental, wild, and cultivated *Prunus* species. We also identified variants of NSPaV in *P. cerasus* and *P. brigantina*, representing new hosts and, in the case of *P. brigantina*, the first report of a wild NSPaV host. GenBank data available to date indicate rather narrow natural host ranges for *Prunus*-infecting luteoviruses. On the other hand, experimental graft inoculations have demonstrated that NSPaV is able to infect *P. tomentosa* and Bing cherry (*P. avium*) indicators (Villamor et al. 2016), suggesting the possibility of a broader natural host range as reported here.

Unlike most other luteoviruses, the genome organization of *Prunus*-infecting luteoviruses shows significant ORF presence/absence variability depending on virus or isolate (ORFs 3a, 4, 6 and 7, Fig. 1A and Table 2). P3a and P4 (MP) have been shown to be involved in luteovirus movement (Ali et al. 2014; Ju et al. 2017; Smirnova et al. 2015). However, these two proteins appear to be dispensable in at least some of the *Prunus*-infecting luteoviruses as already described for NSPaV and AILV1 (Bag et al. 2015; Khalili et al. 2020). Interestingly, we found no evidence for an ORF3a in the PmaLV genome, although it encodes an MP ORF. Despite being the most prevalent luteovirus in peach in our survey, NSPaV lacks both ORF3a and ORF4, both of which are involved in movement. In cases

where it was found in single infection, it could not have been complemented for movement by other viruses, and the mechanism(s) underlying its local and systemic movement therefore remain unclear. The absence of an ORF6 was already known in NSPaV, but was not confirmed in one isolate (sample S7, Table 2). We found no evidence for an ORF6 in PaLV2, while it was present in two isolates of MaLV but absent from another one (sample S3, Table 2). The existence of an ORF7, downstream of ORF6, has been proposed in the case of ChALV (Lenz et al. 2017) and the sequences reported here show that ORF7 is also present in ChLVA. Even if P6 of BYDV-GAV has been shown to have RNA silencing suppression activity in *N. benthamiana* (Liu et al. 2012), the existence of both ORFs 6 and 7 should, however, still be considered speculative since the expression of P6 and P7 *in planta* has yet to be demonstrated (Shen et al. 2006). Altogether, the genomes of *Prunus*-infecting luteoviruses show significant gene composition variation in when it comes to genes involved in RNA silencing suppression and movement. This observation raises questions about possible biological peculiarities of woody *Prunus* hosts and about the strategies used by *Prunus*-infecting luteoviruses to mount systemic invasions of these hosts despite lacking the proteins used to that effect by other luteoviruses.

Perhaps due to their relatively recent discovery, the geographical distribution and prevalence of the *Prunus*-infecting luteoviruses are still poorly known. Obtaining the complete genomes of novel viruses and of additional isolates for known ones has enabled the development of specific diagnostic assays for each of them, allowing us to undertake a systematic survey in European peaches involving 350 samples from eight countries. NSPaV, PaLV, and the novel PaLV2 were identified in each country, a major change in

our vision of the geographic distribution of these viruses. Together with the absence of obvious symptoms and the high prevalence rates observed, these results also suggest that the geographic distribution and prevalence of these viruses may have been largely underestimated and that they are likely present in many other *Prunus*-growing countries. Sequencing of the amplicons generated during the survey indicated that similar to other luteoviruses (Khine et al. 2020; Tian et al. 2019), the genetic diversity of NSPaV, PaLV, and PaLV2 is relatively low. No clustering of isolates based on their geographical origin was identified, a likely consequence of the trade of *Prunus* planting materials and of our inability to detect these agents by widely used biological indexing (Bag et al. 2015).

The results of the retesting of peach trees after two years indicate that these viruses have the ability to persist over extended periods of time in infected *Prunus* hosts. However, PaLV, NSPaV, and PaLV2 were in some cases not re-detected in previously positively tested trees, possibly due to an uneven distribution of infection within host trees. Such a situation is already known for many *Prunus*-infecting viruses (Barba et al. 2011; Büttner et al. 2011; Myrta et al. 2011; Quiot et al. 1995; Salem et al. 2003).

The graft transmissibility of NSPaV had already been demonstrated (Villamor et al. 2016). While confirming these results, the biological indexing experiments performed here on GF305 peach indicator seedlings extend them to PaLV and to the newly identified PaLV2. Surprisingly, graft transmissibility was not 100% for any of these luteoviruses, in contrast to the other co-infecting viruses or viroid. This could be explained by an uneven distribution in the original trees or, alternatively, by another unexpected effect such as the imperfect junction of phloem tissues between the grafted bark pieces and the indicator plants, which might limit transmission of the phloem-limited luteoviruses. It is noteworthy

535 that the virus with the highest graft transmission efficiency, NSPaV, misses ORFs 3a, 4,
536 and 6, which are implicated in viral movement in other luteoviruses, further questioning
537 how *Prunus*-infecting luteoviruses are able to spread in their hosts.

538 Whereas most species of the genus *Luteovirus* are responsible for symptoms and yield
539 reduction (Miller and Lozier, 2022), there are significant uncertainties about the
540 pathogenicity of NSPaV and PaLV. In the present work, none of the analyzed NSPaV,
541 PaLV, or PaLV2 isolates induced clear or reproducible symptoms, alone or in combination,
542 in the widely used GF305 peach indicator. Likewise, detailed symptoms observation of a
543 wide range of orchard-grown peach varieties infected by various combinations of NSPaV,
544 PaLV and PaLV2 failed to identify stem pitting or other unusual symptoms. Taken
545 together, all results reported here suggest an absence of pathogenicity of these viruses
546 in peach under a wide range of situations. Therefore, we suggest that these viruses should
547 likely be considered harmless until proven otherwise in an unambiguous fashion.

548 Another question unanswered to date and with relevance for risk assessment is whether
549 these viruses are transmitted by aphids. Aphids generally transmit luteoviruses in a
550 circulative non-propagative manner (Miller and Lozier, 2022). The mean genetic
551 diversities observed in the BRC orchard for the various viruses are of the same order as
552 their world diversities. This suggests that the observed high infection rates do not result
553 from a local epidemic spread driven by aphids. Similar to AILV1, NSPaV ORF5 is much
554 shorter than in other luteoviruses (Bag et al. 2015; Khalili et al. 2020), while the P3-P5
555 fusion protein is well known to be involved in aphid transmission of luteoviruses (Miller
556 and Lozier, 2022), directly raising the question of NSPaV aphid transmissibility. The

indexing experiments reported here have generated GF305 indicators with single infections, which are excellent starting materials for further aphid transmission studies.

In conclusion, we identified five new luteoviruses from cultivated, wild and ornamental *Prunus* species. We also identified new natural hosts of NSPaV and provided an inclusive and expanded insight into the genetic diversity, geographical distribution, and prevalence of peach-infecting luteoviruses. Taken together, the results obtained point to a lack of pathogenicity of those viruses or to an ability to cause symptoms limited to some specific and possibly infrequent situations. For future research, they also raise interesting questions about the ability of these viruses to mount systemic infections in their *Prunus* hosts despite lacking proteins contributing to the needed functions in other luteoviruses.

Acknowledgments

The authors would like to thank the Platform GetPlaGe (GenoToul, INRAE, Toulouse, France) for the Illumina sequencing, P. Briard and T. Mauduit (UMR BFP, INRAE, Bordeaux) for plant indexing, D. Racofier and M.L. Greil (UE Arboricole, INRAE, Toulenne) for arranging access to *Prunus* BRC, Dr. F. Paprštein (Research And Breeding Fruit Institute, Holovousy, Czech Republic) for arranging access to *Prunus* samples, as well as L. Svanella-Dumas (UMR BFP, INRAE, Bordeaux) for help in collecting samples. The authors declare that they have no conflict of interest.

Literature Cited

Ali, M., Hameed, S., and Tahir, M. 2014. Luteovirus: insights into pathogenicity. Arch. Virol. 159:2853-2860.

Bag, S., Al Rwahnih, M., Li, A., Gonzalez, A., Rowhani, A., Uyemoto, J. K., Sudarshana, M. R. 2015. Detection of a new luteovirus in imported nectarine trees: a case study to propose adoption of metagenomics in post-entry quarantine. Phytopathology 105:840-846.

Barath, D., Jaksa-Czotter, N., Varga, T., and Varallyay, E. 2022. Viromes of Hungarian Peach Trees Identified by High-Throughput Sequencing of Small RNAs. Plants 11:1591.

Barba, M., Hadidi, A., Candresse T., and Cambra M. 2011. *Plum pox virus*. Pages 185-197 in: Virus and virus-like diseases of pome and stone fruits. A. Hadidi, M. Barba, T. Candresse, and W. Jelkmann, eds. American Phytopathological Society Press.

Bester, R., Malan, S. S., and Maree, H. J. 2020. A plum marbling conundrum: identification of a new viroid associated with marbling and corky flesh in Japanese plums. Phytopathology 110:1476-1482.

Büttner, C., von Bargen, S., Bandte, M., and Myrta, A. 2011. Cherry leaf roll virus. Pages 119-125 in: Virus and virus-like diseases of pome and stone fruits. A. Hadidi, M. Barba, T. Candresse, and W. Jelkmann, eds. American Phytopathological Society Press.

Candresse, T., Faure, C., Theil, S., and Marais, A. 2017. First report of nectarine stem pitting-associated virus infecting *Prunus mume* in Japan. Plant Dis. 101:393.

Chang, S., Puryear, J., and Cairney, J. 1993. A simple and efficient method for isolating

599 RNA from pine trees. Plant Mol. Biol. Report 11:113-116.

600 De Paulo, J. J., and Powell, C. A. 1995. Extraction of double-stranded RNA from plant
601 tissues without the use of organic solvents. Plant Dis. 79:246-248.

602 Dinesh-Kumar, S. P., and Miller, W. A. 1993. Control of start codon choice on a plant viral
603 RNA encoding overlapping genes. Plant Cell 5:679-692.

604 Domier, L. L., and D'Arcy, C. J. 2008. Luteoviruses. Pages 231-238 in: Encyclopedia of
605 Virology (Third Edition).B. W. J. Mahy and M. H. V.Van Regenmortel, eds. Elsevier Ltd.

606 Edgar, R. C., Taylor, J., Lin, V., Altman, T., Barbera, P., Meleshko, D., Lohr, D.,
607 Novakovsky, G., Buchfink, B., Al-Shayeb, B., Banfield, J. F., de la Peña, M.,
608 Korobeynikov, A., Chikhi, R., and Babaian, A. 2022. Petabase-scale sequence alignment
609 catalyses viral discovery. Nature 602:142-147

610 Foissac, X., Svanella-Dumas, L., Gentit, P., Dulucq, M. J., Marais, A. and Candresse T.
611 2005. Polyvalent Degenerate Oligonucleotides Reverse Transcription-Polymerase Chain
612 Reaction: A Polyvalent Detection and Characterization Tool for Trichoviruses,
613 Capilloviruses, and Foveaviruses. Phytopathology 95:617-625.

614 François, S., Filloux, D., Fernandez, E., Ogliastro, M., and Roumagnac, P. 2018. Viral
615 metagenomics approaches for high-resolution screening of multiplexed arthropod and
616 plant viral communities. Methods Mol Biol. 1746:77-95.

617 Hadidi, A., and Barba, M. 2011. Economic impact of pome and stone fruit viruses and
618 viroids. Pages 1-7 in: Virus and virus-like diseases of pome and stone fruits. A. Hadidi, M.
619 Barba, T. Candresse, and W. Jelkmann, eds. American Phytopathological Society Press.

Hillman, B. I. and Esteban, R. 2011. Family *Luteoviridae.* Pages 1045-1060 in: A. M. Q. King, M. J. Adams, E. B. Carstens, E. J. Lefkowitz, eds. Virus Taxonomy, Ninth Report of International Committee on Taxonomy of Viruses. Amsterdam, The Netherlands, Elsevier Academic Press.

Hou, W., Li, S., and Massart, S. 2020. Is there a “biological desert” with the discovery of new plant viruses? A retrospective analysis for new fruit tree viruses. Front. Microbiol.11:592816.

Igori, D., Baek, D., Kim, S. Y., Seo, E., Lee, S. H., Jeong, R. D., Yi, S. Y., Bong, J. J. and Moon, J. S. 2017a. Complete genome sequence of nectarine stem pitting-associated virus, isolated from *Prunus persica* in Cheongdo County, South Korea. Genome Announc. 5:e00908-17.

Igori, D., Lim, S., Baek, D., Cho, I. S., and Moon, J. S. 2017b. Complete nucleotide sequence of a highly divergent cherry-associated luteovirus (ChALV) isolate from peach in South Korea. Arch. Virol. 162:2893-2896.

Jo, Y., Cho, J. K., Choi, H., Lian, S., and Cho, W. K. 2017. First report of nectarine stem pitting-associated virus and Plum bark necrosis and stem pitting-associated virus infecting a peach cultivar in Korea. Plant Dis. 101:1067.

Ju, J., Kim, K., Lee, K. J., Lee, W. H., and Ju, H. J. 2017. Localization of Barley yellow dwarf virus movement protein modulating programmed cell death in *Nicotiana benthamiana*. Plant Pathol. J. 33:53-65.

Kelly, L., Gerlach, W. L., and Waterhouse, P. M. 1994. Characterisation of the subgenomic RNAs of an Australian isolate of barley yellow dwarf luteovirus. Virology 202:565–573.

642 Khalili, M., Candresse, T., Faure, C., and Marais, A. 2020. Complete genome sequence
643 of almond luteovirus 1, a novel luteovirus infecting almond. Arch. Virol. 165:2123-2126.

644 Khine, M. O., Michaela, B., Yan, L. I. U., Kundu, J. K., and Wang, X. 2020. Molecular
645 diversity of barley yellow dwarf virus-PAV from China and the Czech Republic. J. Integr.
646 Agric. 19:2736-2745.

647 Krizbai, L., Kriston, E., Kreuze, J., and Melika, G. 2017. Identification of nectarine stem
648 pitting-associated virus infecting *Prunus persica* in Hungary. New Dis. Reports 35:588-
649 2044.

650 Kumar, S., Stecher, G., and Tamura, K. 2016. MEGA7: molecular evolutionary genetics
651 analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33:1870-1874.

652 Lenz, O., Přibylová, J., Fránová, J., Koloniuk, I., and Špak, J. 2017. Identification and
653 characterization of a new member of the genus *Luteovirus* from cherry. Arch. Virol.
654 162:587-590.

655 Liu, Y., Zhai, H., Zhao, K., Wu, B., and Wang, X. 2012. Two suppressors of RNA silencing
656 encoded by cereal-infecting members of the family *Luteoviridae*. J. Gen. Virol. 93:1825-
657 1830.

658 Liu, H., Wu, L., Nikolaeva, E., Peter, K., Liu, Z., Mollov, D., Cao, M., Li, R. 2018.
659 Characterization of a new apple luteovirus identified by high-throughput sequencing. Virol.
660 J. 15:1-9.

661 Lu, M. G., Zhang, C., Zhang, Z. X., Wang, C. A., and Li, S. F. 2017. Nectarine stem pitting-
662 associated virus detected in peach trees in China. Plant Dis. 101:513.

Maliogka, V. I., Minafra, A., Saldarelli, P., Ruiz-García, A. B., Glasa, M., Katis, N., and Olmos, A. 2018. Recent Advances on Detection and Characterization of Fruit Tree Viruses Using High-Throughput Sequencing Technologies. Viruses 10:436.

Marais, A., Faure, C., Bergey, B., and Candresse, T. 2018. Viral double-stranded RNAs (dsRNAs) from plants: Alternative nucleic acid substrates for high-throughput sequencing. Methods Mol. Biol. 1746:45-53.

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., and Muhire, B. 2015. RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evol. 1:vev003

Massart, S., Candresse, T., Gil, J., Lacomme, C., Predajna, L., Ravnikar, M., Reynard, J. S., Rumbou, A., Saldarelli, P., Škoric, D., Vainio, E. J., Valkonen, J. P. T., Vanderschuren, H., Varveri, C., and Wetzel, T. 2017. A framework for the evaluation of biosecurity, commercial, regulatory, and scientific impacts of plant viruses and viroids identified by NGS technologies. Front. Microbiol. 8:45.

Miller, W. A., and Rasochová, L. 1997. Barley yellow dwarf viruses. Annu. Rev. Phytopathol. 35:167-190.

Miller, W. A., and Lozier, Z. 2022. Yellow Dwarf Viruses of Cereals: Taxonomy and Molecular Mechanisms. Annu. Rev. Phytopathol. 60:6.1-6.21

Myrta, A., Matic, S., Malinowski, T., Pasquini, G., and Candresse, T. 2011. *Apple chlorotic leaf spot virus* in stone fruits. Pages 85-90 in: Virus and virus-like diseases of pome and stone fruits. A. Hadidi, M. Barba, T. Candresse, and W. Jelkmann, eds. American Phytopathological Society Press.

Quiot, J. B., Labonne, G., Boeglin, M., Adamolle, C., Renaud, L. Y., and Candresse, T.

1995. Behaviour of two isolates of Plum pox virus inoculated on peach and apricot trees: first results. Acta Hortic. 386:290-297.Rubio, M., Martínez-Gómez, P., Marais, A., Sánchez-Navarro, J. A., Pallás, V., and Candresse, T. 2017. Recent advances and prospects in *Prunus* virology. Ann. Appl. Biol. 171:125-138.

Salem, N., Mansour, A., Al-Musa, A., Al-Nsour, A. 2003. Seasonal variation of Prunus necrotic ringspot virus concentration in almond, peach, and plum cultivars. Phytopathol. Mediterr. 42:155-160.

Shen, P., Tian, X., Zhang, S., Ren, F., Li, P., Yu, Y. Q., Li R., Zhou, C., Cao, M. 2018. Molecular characterization of a novel luteovirus infecting apple by next-generation sequencing. Arch. Virol. 163:761-765.

Shen, R., Rakotondrafara, A. M., and Miller, W. A. 2006. trans regulation of cap-independent translation by a viral subgenomic RNA. J. Virol. 80:10045-10054.

Smirnova, E., Firth, A. E., Miller, W. A., Scheidecker, D., Brault, V., Reinbold, C., Rakotondrafara, A. M., Chung, B. Y. W., Ziegler-Graff, V. 2015. Discovery of a small non-AUG-initiated ORF in poleroviruses and luteoviruses that is required for long-distance movement. PLoS Pathog. 11:e1004868.

Sorrentino, R., Marais, A., Faure, C., Theil, S., Alioto, D., and Candresse, T. 2018. First report of peach-associated luteovirus in nectarine (*Prunus persica*) in Italy. Plant Dis. 102:1465.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673-

4680.

Tian, B., Gildow, F. E., Stone, A. L., Sherman, D. J., Damsteegt, V. D., and Schneider, W. L. 2019. Aphid vectors impose a major bottleneck on Soybean dwarf virus populations for horizontal transmission in soybean. Phytopathol. Res. 1:29.

Umer, M., Liu, J., You, H., Xu, C., Dong, K., Luo, N., Kong, L., Li, X., Hong, N., Wang, G., Fan, X., Kotta-Loizou, I., Xu W. 2019. Genomic, Morphological and Biological Traits of the Viruses Infecting Major Fruit Trees. Viruses 11:515.

Villamor, D. E. V, Mekuria, T. A., Pillai, S. S., and Eastwell, K. C. 2016. High-throughput sequencing identifies novel viruses in nectarine: insights to the etiology of stem-pitting disease. Phytopathology 106:519-527.

Walls, J., Rajotte, E., and Rosa, C. 2019. The past, present, and future of barley yellow dwarf management. Agriculture 9:23.

Wu, L.-P., Liu, H.-W., Bateman, M., Liu, Z., and Li, R. 2017. Molecular characterization of a novel luteovirus from peach identified by high-throughput sequencing. Arch. Virol. 162:2903-2905.

Zhou, J., Zhang, Z., Lu, M., Xiao, H., and Li, S. 2018. First report of peach-associated luteovirus from flat peach and nectarine in China. Plant Dis. 102:2669.

TABLES

TABLE 1. List of *Prunus* samples from which luteovirus genomes were reconstructed in the present work

Index name	Species	Variety / Cultivar	Nature/type	Symptoms	Collecting location	Country of origin	Collection year
S1	*Prunus mume*	not known	Ornamental	Oak leaf mosaic	Kyoto, Japan	Japan	2015
S2	*Prunus armeniaca*	Jia Na Li	Cultivated	Mosaics, leaf and twig deformation	Germplasm [a], Czech Republic	China	2021
S3	*Prunus incisa*	na	Wild	No	Germplasm [b], France	Japan	2019
S4	*Prunus persica*	Henri Moulin	Cultivated	No	Germplasm [b], France	France	2019
S5	*Prunus cerasus*	Rannaja	Cultivated	No	Czech Republic	Moldova	2015
S6	*Prunus cerasus*	Cigany	Cultivated	Mosaic	Czech Republic	Hungary	2015
S7	*Prunus cerasus*	Amarelka Chvalkovicka	Cultivated	No	Germplasm [c], Czech Republic	Czech Republic	2013
S8	*Prunus mahaleb*	na	Wild	Bushy growth and shortened internodes	Aussois, France	France	2021
S9	*Prunus brigantina*	na	Wild	Bushy growth and shortened internodes	Névache mountain, France	France	2017

[a]: Mendel University (Mendelu Lednice, Czech Republic), [b]: *Prunus* INRAE Biological Resource Center (BRC Toulenne, France), [c]: Research and Breeding Institute of Pomology (VŠÚO, Holovousy, Czech Republic), na: not applicable

TABLE 2. Molecular features of representative *Prunus*-infecting luteoviruses

Virus	Genome size (nt)	Protein size (aa)								Reference
		P1	RdRp-fusion	P3a	P3 (CP)	CP-RTD	MP	P6	P7	
ChLVA-Rannaja	5,726	364	890	47	196	630	170	37	71	This study
MaLV-*P. mume*	5,748	368	895	48	197	642	175	74	na	This study
PaLV2	5,780	364	890	48	195	640	170	na	na	This study
PmaLV	5,822 [a]	364	890	Na	198	647	147	38	na	This study
PhaLV	5,202 [a]	364	890	48	196	632	172	50	na	This study
NSPaV-*P. cerasus*	4,993 [a]	328	847	Na	206	526	na	62	na	This study
PaLV-NC034970	5,819	364	890	49	199	670	177	56	49	Wu et al. 2017
ChALV-NC031800	5,857	364	890	45	198	647	175	79	79	Lenz et al. 2017
AlLV1-MT362517	5,047	329	848	Na	204	550	na	na	na	Khalili et al. 2020
NSPaV-NC027211	4,991	328	847	Na	206	526	na	na	na	Bag et al. 2015

[a] = not completed by Race experiments; na: not applicable; NSPaV: nectarine stem pitting-associated virus; PmaLV: Prunus mahaleb-associated luteovirus; PhaLV: Prunus humilis-associated luteovirus; PaLV2: peach-associated luteovirus 2; MaLV: mume-associated luteovirus; ChLVA: cherry luteovirus A; PaLV: peach-associated luteovirus; ChALV: cherry-associated luteovirus; AlLV1: almond luteovirus 1; RdRp: RNA-dependent RNA polymerase; CP: Coat protein; CP-RTD: CP-readthrough domain; MP: Movement protein.

TABLE 3. Peach-infecting luteovirus incidence in *Prunus persica* in eight European countries

Geographical origin	**Number of samples**	**Number of collection sites**	**Peach infecting viruses**		
			NSPaV	**PaLV**	**PaLV2**
Belgium	26	1 germplasm	53%	57%	15%
Czech Republic	43	1 germplasm + 6 orchards	100%	60%	9%
France	50	1 germplasm	96%	38%	54%
Greece	30	5 orchards	53%	13%	3%
Italy	51	1 germplasm	27%	26%	4%
Slovakia	50	1 germplasm + 5 orchards	88%	88%	6%
Spain	50	44 orchards	68%	30%	10%
Turkey	50	4 orchards	42%	6%	8%

NSPaV: nectarine stem pitting-associated virus; PaLV: peach-associated luteovirus; PaLV2: peach-associated luteovirus 2

TABLE 4. Graft transmission experiments of NSPaV, PaLV and PaLV2 on GF305 peach seedling indicator plants

Peach accession	**Infection status**	**Biological indexing**		**Luteovirus transmission**			**Other viruses/viroids transmission**		
		Symptoms [a]		Positive / Grafted			Positive / Grafted		
		Cycle 1	Cycle 2	PaLV2	NSPaV	PaLV	PNRSV	ACLSV	PLMVd
S2686 X5Y70 O3	PaLV2-NSPaV-PNRSV	AS, 10/10	AS, 10/10	1/10	8/10	na	10/10	na	na
S4072 X12Y24 Q	PaLV2-NSPaV	AS, 7/7	AS, 7/7	2/7	6/7	na	na	na	na
S3527 X2Y16 O1	PaLV2-NSPaV	LC, LR, 2/9 □	LC, LR, 2/9	0/9	3/9	na	na	na	na
S5555 X4Y67 O2	PaLV2-NSPaV-PaLV	AS, 10/10	LC, LR, 1/10	0/10	2/10	1/10	na	na	na
S2464 X1Y16 Q	NSPaV-PaLV	AS, 10/10	AS, 10/10	Na	7/10	3/10	na	na	na
S4617 X2Y45 O2	NSPaV-PaLV	AS, 4/4	AS, 4/4	Na	3/4	0/4	na	na	na
S2464 X5Y76 O3	NSPaV-PaLV	AS, 10/10	AS, 10/10	Na	5/10	4/10	na	Na	na
S1932 X1Y7 Q	PaLV-PNRSV-ACLSV-PLMVd	DGSM, 6/6	DGSM, 6/6	Na	na	4/6	6/6	6/6	6/6
S1161 X7Y8 O3	NSPaV	AS, 5/5	SN, 1/5	Na	4/5	na	na	Na	na
S8278 X6Y75 O3	NSPaV	LD, 3/9	AS, 9/9	na	3/9	na	na	Na	na
Overall transmission rate				8.3%	55.4%	30%	100%	100%	100%

[a] AS: asymptomatic, LC: leaf chlorosis, LR: leaf reddening, □: decline and death, DGSM: dark green sunken mottle, LD: leaf deformation, SN: stem necrosis

na: does not apply; NSPaV: nectarine stem pitting-associated virus; PaLV2: peach-associated luteovirus 2; PaLV: peach-associated luteovirus; ACLSV: apple chlorotic leaf spot virus; PNRSV: prunus necrotic ringspot virus; PLMVd: peach latent mosaic viroid

FIGURE CAPTIONS

Fig. 1. Genomic organization *Prunus*-infecting luteoviruses (A) and phylogenetic tree based on their whole genome sequence alignment (B). The newly discovered viruses in this study are shown by triangles and the divergent variants by circles. The phylogenetic tree was constructed using the neighbor joining method in MEGA7 and a strict nucleotide identity distance. Bootstrap values (1,000 replicates) less than 70% were removed. PmaLV: Prunus mahaleb-associated luteovirus; ChALV: cherry-associated luteovirus; PaLV: peach-associated luteovirus; PhaLV: Prunus humilis-associated luteovirus; ChLVA: cherry luteovirus A; PaLV2: peach-associated luteovirus 2; MaLV: mume associated luteovirus; AlLV1: almond luteovirus 1; NSPaV: nectarine stem pitting-associated virus. The scale bar represents 5% nucleotide divergence. ORF1: open reading frame 1 Pol: RNA-dependent RNA polymerase; MP: movement protein; CP: coat protein; RT: readthrough domain.

Fig. 2. Phylogenetic trees based on the alignment of the nucleotide sequences of the luteoviral PCR products generated from the positive samples from different countries. A. Nectarine stem pitting-associated virus. B. Peach-associated luteovirus. C. Peach-associated luteovirus 2. GenBank reference sequences are indicated by black dots. The geographical origin of the isolates is summarized as follows: SP: Spain; Tr: Turkey; Bl: Belgium; Gr: Greece; Fr: France; Cz: Czech Republic; IT: Italy; Sk: Slovakia. The phylogenetic trees were constructed using neighbor joining method in MEGA7 and strict nucleotide identity distances. Bootstrap values (1,000 replicates) less than 70% are not shown. The scale bars represent 0.5% (A and C) or 1% nucleotide divergence (B).

SUPPLEMENTARY FIGURE CAPTIONS

Supplementary Fig. S1. Phylogenetic trees based on the alignment of the P1-P2 (A) and P3-P5 (B) aa deduced sequences of *Prunus*-infecting luteoviruses. Phylogenetic trees were constructed using the neighbor joining method in MEGA7 and a strict aa identity distance. Bootstrap values (1,000 replicates) less than 70% were removed. The scale bars represent 5% aa divergence

Supplementary Fig. S2. Pairwise aa identity of *Prunus*-infecting luteoviruses in P1-P2 (CP-RTD) fusion protein (**A**) and in P3-P5 (**B**) (viral replicase)

SUPPLEMENTARY TABLE TITLES

Supplementary Table S1. List of primers used in this study

Supplementary Table S2. Methods used for HTS, number of trimmed reads, average coverage and mapped reads percent to the reference genome of novel *Prunus*-infecting luteoviruses and new isolates.

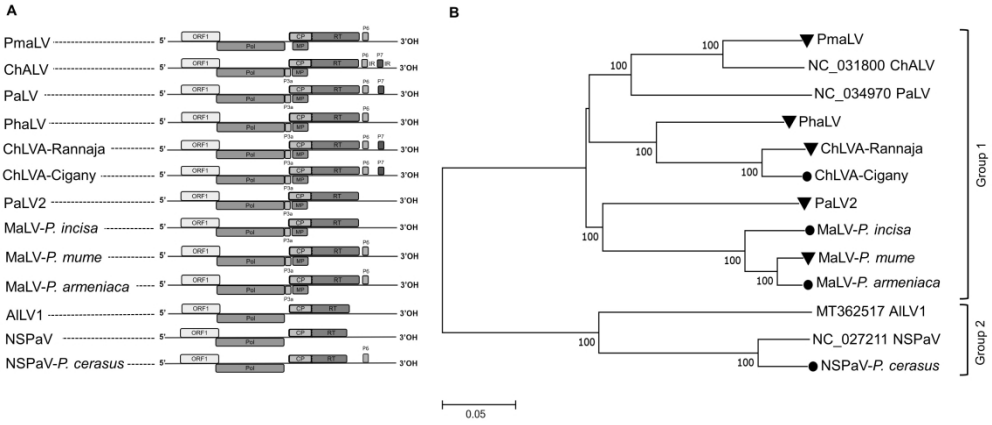
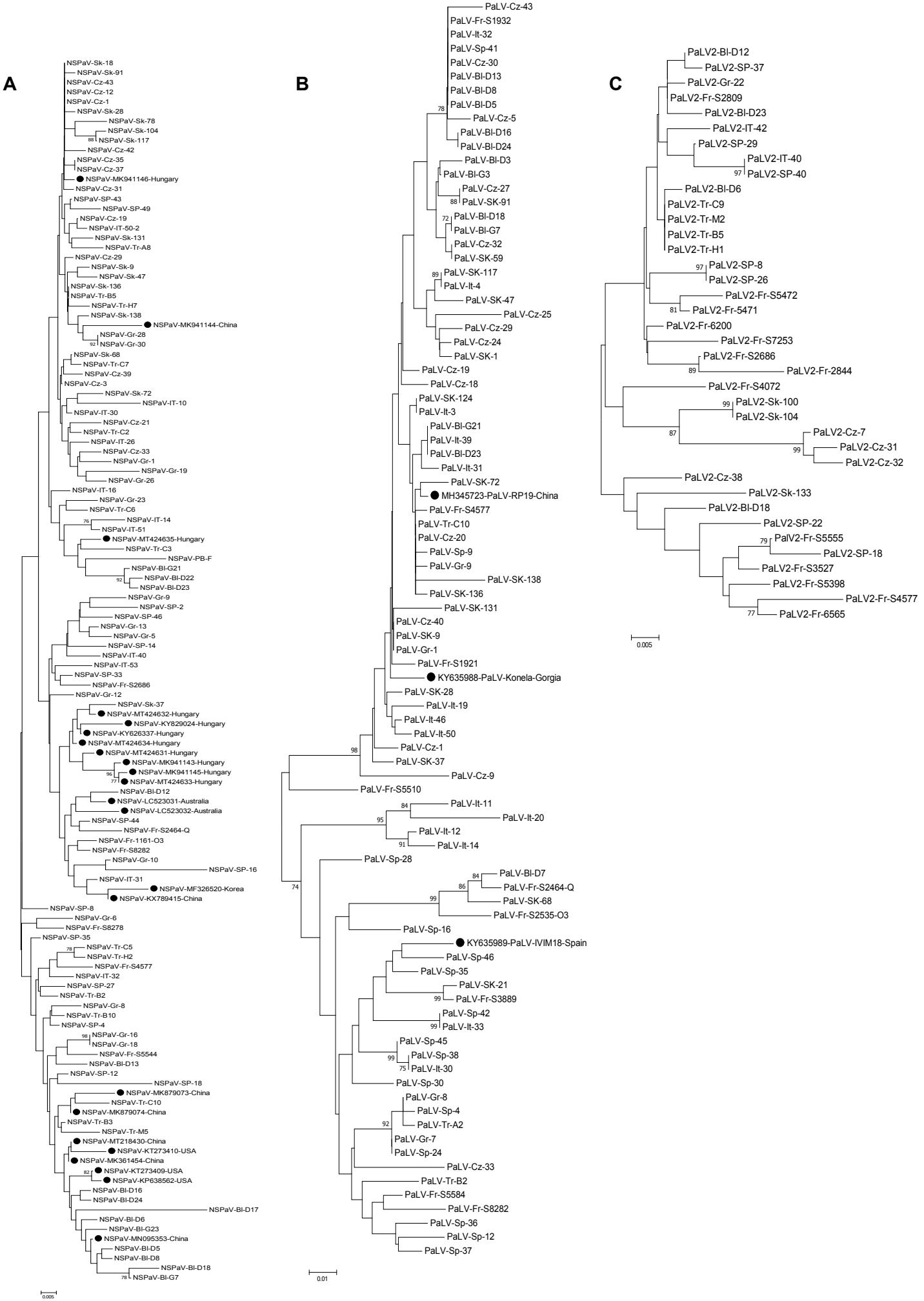


Fig. 1. Genomic organization Prunus-infecting luteoviruses (A) and phylogenetic tree based on their whole genome sequence alignment (B). The newly discovered viruses in this study are shown by triangles and the divergent variants by circles. The phylogenetic tree was constructed using the neighbor joining method in MEGA7 and a strict nucleotide identity distance. Bootstrap values (1,000 replicates) less than 70% were removed. PmaLV: Prunus mahaleb-associated luteovirus; ChALV: cherry-associated luteovirus; PaLV: peach-associated luteovirus; PhaLV: Prunus humilis-associated luteovirus; ChLVA: cherry luteovirus A; PaLV2: peach-associated luteovirus 2; MaLV: mume associated luteovirus; AlLV1: almond luteovirus 1; NSPaV: nectarine stem pitting-associated virus. The scale bar represents 5% nucleotide divergence. ORF1: open reading frame 1 Pol: RNA-dependent RNA polymerase; MP: movement protein; CP: coat protein; RT: readthrough domain.

338x141mm (300 x 300 DPI)


A
NSPaV-Sk-18
NSPaV-Sk-91
NSPaV-Cz-43
NSPaV-Cz-12
NSPaV-Cz-1
NSPaV-Sk-28
NSPaV-Sk-78
NSPaV-Sk-104
NSPaV-Sk-117
NSPaV-Cz-42
NSPaV-Cz-35
NSPaV-Cz-37
NSPaV-MK941146-Hungary
NSPaV-Cz-31
NSPaV-SP-43
NSPaV-SP-49
NSPaV-Cz-19
NSPaV-IT-50-2
NSPaV-Sk-131
NSPaV-Tr-A8
NSPaV-Cz-29
NSPaV-Sk-9
NSPaV-Sk-47
NSPaV-Sk-136
NSPaV-Tr-B5
NSPaV-Tr-H7
NSPaV-Sk-138
NSPaV-MK941144-China
NSPaV-Gr-28
NSPaV-Gr-30
NSPaV-Sk-68
NSPaV-Tr-C7
NSPaV-Cz-39
NSPaV-Cz-3
NSPaV-Sk-72
NSPaV-IT-10
NSPaV-IT-30
NSPaV-Cz-21
NSPaV-Tr-C2
NSPaV-IT-26
NSPaV-Cz-33
NSPaV-Gr-1
NSPaV-Gr-19
NSPaV-Gr-26
NSPaV-IT-16
NSPaV-Gr-23
NSPaV-Tr-C6
NSPaV-IT-14
NSPaV-IT-51
NSPaV-MT424635-Hungary
NSPaV-Tr-C3
NSPaV-PB-F
NSPaV-Bl-G21
NSPaV-Bl-D22
NSPaV-Bl-D23
NSPaV-Gr-9
NSPaV-SP-2
NSPaV-SP-46
NSPaV-Gr-13
NSPaV-Gr-5
NSPaV-SP-14
NSPaV-IT-40
NSPaV-IT-53
NSPaV-SP-33
NSPaV-Fr-S2686
NSPaV-Gr-12
NSPaV-Sk-37
NSPaV-MT424632-Hungary
NSPaV-KY829024-Hungary
NSPaV-KY626337-Hungary
NSPaV-MT424634-Hungary
NSPaV-MT424631-Hungary
NSPaV-MK941143-Hungary
NSPaV-MK941145-Hungary
NSPaV-MT424633-Hungary
NSPaV-Bl-D12
NSPaV-LC523031-Australia
NSPaV-LC523032-Australia
NSPaV-SP-44
NSPaV-Fr-S2464-Q
NSPaV-Fr-1161-O3
NSPaV-Fr-S8282
NSPaV-Gr-10
NSPaV-SP-16
NSPaV-IT-31
NSPaV-MF326520-Korea
NSPaV-KX789415-China
NSPaV-SP-8
NSPaV-Gr-6
NSPaV-Fr-S8278
NSPaV-SP-35
NSPaV-Tr-C5
NSPaV-Tr-H2
NSPaV-Fr-S4577
NSPaV-IT-32
NSPaV-SP-27
NSPaV-Tr-B2
NSPaV-Gr-8
NSPaV-Tr-B10
NSPaV-SP-4
NSPaV-Gr-16
NSPaV-Gr-18
NSPaV-Fr-S5544
NSPaV-Bl-D13
NSPaV-SP-12
NSPaV-SP-18
NSPaV-MK879073-China
NSPaV-Tr-C10
NSPaV-MK879074-China
NSPaV-Tr-B3
NSPaV-Tr-M5
NSPaV-MT218430-China
NSPaV-KT273410-USA
NSPaV-MK361454-China
NSPaV-KT273409-USA
NSPaV-KP638562-USA
NSPaV-Bl-D16
NSPaV-Bl-D24
NSPaV-Bl-D17
NSPaV-Bl-D6
NSPaV-Bl-G23
NSPaV-MN095353-China
NSPaV-Bl-D5
NSPaV-Bl-D8
NSPaV-Bl-D18
NSPaV-Bl-G7
0.005
B
PaLV-Cz-43
PaLV-Fr-S1932
PaLV-It-32
PaLV-Sp-41
PaLV-Cz-30
PaLV-Bl-D13
PaLV-Bl-D8
PaLV-Bl-D5
PaLV-Cz-5
PaLV-Bl-D16
PaLV-Bl-D24
PaLV-Bl-D3
PaLV-Bl-G3
PaLV-Cz-27
PaLV-SK-91
PaLV-Bl-D18
PaLV-Bl-G7
PaLV-Cz-32
PaLV-SK-59
PaLV-SK-117
PaLV-It-4
PaLV-SK-47
PaLV-Cz-25
PaLV-Cz-29
PaLV-Cz-24
PaLV-SK-1
PaLV-Cz-19
PaLV-Cz-18
PaLV-SK-124
PaLV-It-3
PaLV-Bl-G21
PaLV-It-39
PaLV-Bl-D23
PaLV-It-31
PaLV-SK-72
MH345723-PaLV-RP19-China
PaLV-Fr-S4577
PaLV-Tr-C10
PaLV-Cz-20
PaLV-Sp-9
PaLV-Gr-9
PaLV-SK-138
PaLV-SK-136
PaLV-SK-131
PaLV-Cz-40
PaLV-SK-9
PaLV-Gr-1
PaLV-Fr-S1921
KY635988-PaLV-Konela-Gorgia
PaLV-SK-28
PaLV-It-19
PaLV-It-46
PaLV-It-50
PaLV-Cz-1
PaLV-SK-37
PaLV-Cz-9
PaLV-Fr-S5510
PaLV-It-11
PaLV-It-20
PaLV-It-12
PaLV-It-14
PaLV-Sp-28
PaLV-Bl-D7
PaLV-Fr-S2464-Q
PaLV-SK-68
PaLV-Fr-S2535-O3
PaLV-Sp-16
KY635989-PaLV-IVIM18-Spain
PaLV-Sp-46
PaLV-Sp-35
PaLV-SK-21
PaLV-Fr-S3889
PaLV-Sp-42
PaLV-It-33
PaLV-Sp-45
PaLV-Sp-38
PaLV-It-30
PaLV-Sp-30
PaLV-Gr-8
PaLV-Sp-4
PaLV-Tr-A2
PaLV-Gr-7
PaLV-Sp-24
PaLV-Cz-33
PaLV-Tr-B2
PaLV-Fr-S5584
PaLV-Fr-S8282
PaLV-Sp-36
PaLV-Sp-12
PaLV-Sp-37
0.01
C
PaLV2-Bl-D12
PaLV2-SP-37
PaLV2-Gr-22
PaLV2-Fr-S2809
PaLV2-Bl-D23
PaLV2-IT-42
PaLV2-SP-29
PaLV2-IT-40
PaLV2-SP-40
PaLV2-Bl-D6
PaLV2-Tr-C9
PaLV2-Tr-M2
PaLV2-Tr-B5
PaLV2-Tr-H1
PaLV2-SP-8
PaLV2-SP-26
PaLV2-Fr-S5472
PaLV2-Fr-5471
PaLV2-Fr-6200
PaLV2-Fr-S7253
PaLV2-Fr-S2686
PaLV2-Fr-2844
PaLV2-Fr-S4072
PaLV2-Sk-100
PaLV2-Sk-104
PaLV2-Cz-7
PaLV2-Cz-31
PaLV2-Cz-32
PaLV2-Cz-38
PaLV2-Sk-133
PaLV2-Bl-D18
PaLV2-SP-22
PalV2-Fr-S5555
PaLV2-SP-18
PaLV2-Fr-S3527
PaLV2-Fr-S5398
PaLV2-Fr-S4577
PaLV2-Fr-6565
0.005

Supplementary Table S1. List of primers used in this study

Primer name	Use	Sequence 5'-3'	Annealing temperature	PCR product size (nt)
MaLV-F	Detection	TCTACGAAGGATGATCAGTTCAA	55°C	549
MaLV-R	Detection	GAACAATTTGAATAGTTCCCTA		
MaLV-5RACE[a]	5' RACE	ACTCGAAGCGTAGATGAGCGAATC	70°C	150
MaLV-3RACE [a]	3' RACE	CTACCTAGTCAGGGGGATGGCTCACCATGTT	70°C	389
PaLV-F2	Detection	CTTTGGCGGCTAGGGCTTGCA	60°C	282
PaLV-R2	Detection	GAGAAGAGCCTCCGCTACCATTTA		
PaLV2-F	Detection	AGTCAGGTAGACGTCGTTGTAAA	61°C	365
PaLV2-R	Detection	TCTTCGGTGGTGCCCTCATTCTC		
PaLV2-5RACE[a]	5' RACE	GTGCCCTCATTCTCCCCTCCCTTGACCT	70°C	881
PaLV2-3RACE[a]	3' RACE	GTGGTGGACTATCGTTGTGAGGTGTG	70°C	475
PhaLV-F	Detection	GTCCTCCATATCGTGAAGAGA	56°C	308
PhaLV-R	Detection	AAGCGGGTTGGACTTTGCTGT		
ChLVA-1492	5' RACE	ACGTTGGTATATAGGTATGACAC	60°C	191
ChLVA-1493	5' RACE	GCATTCCCATTCCCATTCTT	60°C	318
ChLVA-1494	3' RACE	AATTGGTAGTTCTGTTGTCA	60°C	530
ChLVA-1495	3' RACE	TTACGTGTTAGTTGAAGGTT	60°C	415
ChLVA-1684	3' RACE	TGGTCACCTCGTTAAACAAC	60°C	487
NSPaV-F2	Detection	ACGACAAGGCGCACCCGCACCTC	62°C	335
NSPaV-R2	Detection	TCTGGGTGCAACTAGTGTCAATC		
LDF-087	S3 cDNA synthesis	TATGCTCGACCGCCNNNNNNNNNNNT	42°C	na
Tag718	S3 PCR for HTS	GGTCTTACATTATGCTCGACCGCC	45°C	na
LDF-042	S4 cDNA synthesis	CACTGAGCACCCCGGTCGCTATCA	42°C	na
Tag782	S4 PCR for HTS	CCCGGTCGCTATCANNNNNNNNNNNT	45°C	na
PcDNA12	S1, S7 cDNA synthesis	TTGGGTGTGTTTGGNNNNNNNNNNNT	42°C	na
MID-GENCO14	S1 PCR for HTS	CAAGAGTTTGTGTTGGGTGTGTTTGG	65°C-45°C	na
MID-GENCO6	S7 PCR for HTS	AGAGTCTTTGTGTTGGGTGTGTTTGG	65°C-45°C	na

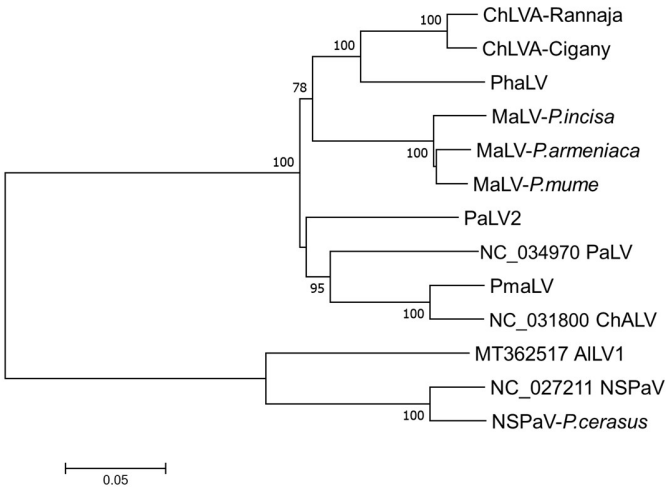
This primer was used in conjunction with the universal primer provided by the 5′ and 3' rapid amplification of cDNA ends kit (Takara Bio Europe)

Supplementary Table S2. Methods used for HTS, number of trimmed reads, average coverage and mapped reads percent to the reference genome of novel *Prunus*-infecting luteoviruses and new isolates.

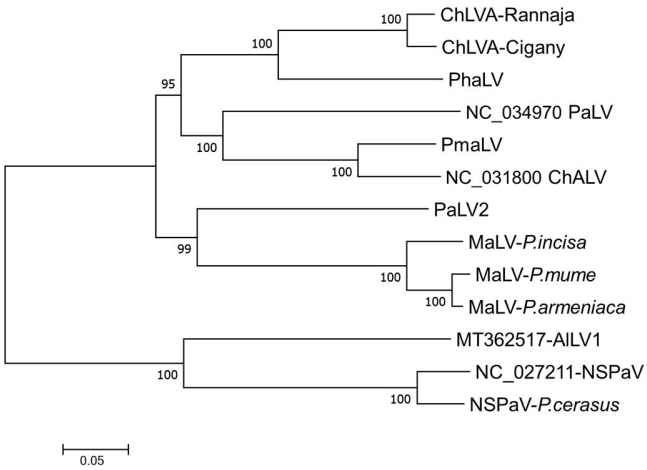
Sample	Virus-isolate	Method	Trimmed reads	Average coverage	mapped reads (%)	Accession numbers
S1	MaLV-*P. mume*	dsRNA	33,113	18.6x	1.54	ON408236
S2	MaLV-*P. armeniaca*	RNA	122,714,448	40.8x	0.0015	ON408237
S3	MaLV-*P. incisa*	dsRNA	415,410	855.7x	1.18	ON408235
S4	PaLV2	dsRNA	945,443	128.22x	0.69	ON408234
S5	ChLVA-Rannaja	dsRNA	25,890,504	29.7x	0.007	ON146357
S6	ChLVA-Cigany	dsRNA	25,944,878	20.2x	0.005	ON146356
S7	NSPaV-*P. cerasus*	dsRNA	720,762	1x	0.01	ON408233
S8	PmaLV	RNA	58,149,924	155.45x	0.01	ON408238
na	PhaLV	Datamining	124,143,971	51.6x	0.002	BK061315

na: not applicable; MaLV: mume-associated luteovirus; PaLV2: peach-associated luteovirus 2; ChLVA: cherry luteovirus A; NSPaV: nectarine stem pitting-associated virus; PmaLV: prunus mahaleb-associated luteovirus; PhaLV: prunus humilis-associated luteovirus

A



B



<u>Supplementary Fig. S1.</u> Phylogenetic trees based on the alignment of the P1-P2 (A) and P3-P5 (B) aa deduced sequences of *Prunus* infecting luteoviruses. Phylogenetic trees were constructed using the neighbor joining method in MEGA7 and a strict aa identity distance. Bootstrap values (1,000 replicates) less than 70% were removed. The scale bars represent 5% aa divergence

A

MaLV-*P.mume*	100																
MaLV-*P.armeniaca*	97	100															
MaLV-*P.incisa*	96	96	100														
PaLV2-*P.persica*	84	84	83	100													
ChLVA-Rannaja	84	84	83	84	100												
ChLVA-Cigany	85	84	84	84	97	100											
PhaLV	83	84	83	83	88	88	100										
PmaLV	82	81	81	83	82	82	82	100									
NSPaV-*P.cerasus*	52	53	52	53	52	53	53	53	100								
NC_031800-ChALV	82	82	81	84	82	82	81	95	52	100							
NC_034970-PaLV	84	83	83	84	82	82	81	85	52	85	100						
NC_027211-NSPaV	52	53	52	54	52	53	53	54	94	52	53	100					
MT362517-AlLV1	53	53	53	54	53	53	53	54	79	53	53	79	100				
NC_040680-ALV1	63	63	62	62	61	62	62	64	53	63	63	53	51	100			
NC_040549-AaLV	62	62	61	63	61	61	62	64	53	63	63	53	53	64	100		
NC_010806-RSDaV	55	55	55	56	55	55	56	56	51	57	55	51	52	53	55	100	
NC_004750-BYDV-PAV	52	52	52	52	52	52	52	53	49	53	51	49	50	51	51	51	100

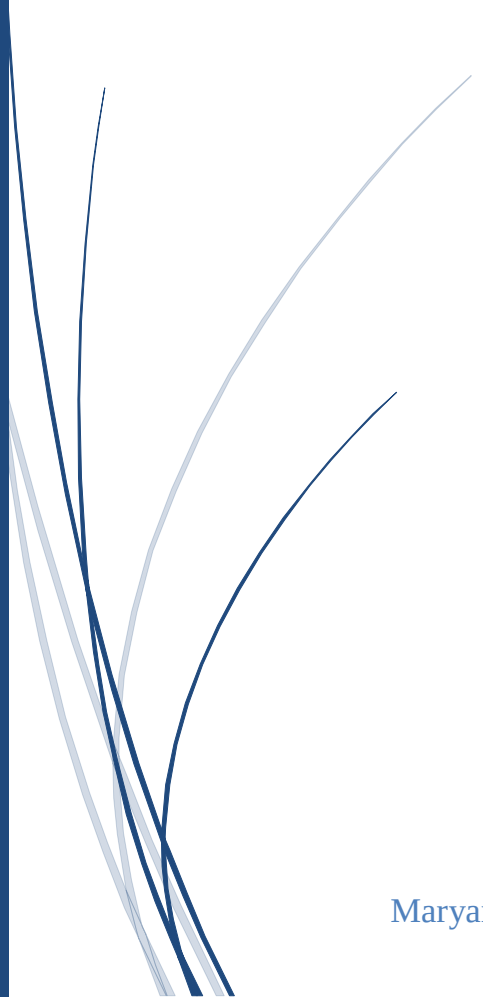
B

MaLV-*P.mume*	100																
MaLV-*P.armeniaca*	98	100															
MaLV-*P.incisa*	90	92	100														
PaLV2-*P.persica*	62	62	62	100													
ChLVA-Rannaja	54	54	55	60	100												
ChLVA-Cigany	54	55	55	59	96	100											
PhaLV	53	54	54	58	76	76	100										
PmaLV	54	55	54	58	61	61	62	100									
NSPaV-*P.cerasus*	30	30	30	31	30	29	31	31	100								
NC_031800-ChALV	54	55	54	57	61	61	59	88	31	100							
NC_034970-PaLV	52	52	52	55	58	58	59	65	31	65	100						
NC_027211-NSPaV	30	30	29	30	29	29	30	30	92	30	30	100					
MT362517-AlLV1	30	30	31	33	31	31	32	32	58	32	33	58	100				
NC_040680-ALV1	36	36	35	37	37	37	36	38	29	37	35	29	30	100			
NC_040549-AaLV	38	38	37	37	39	39	39	38	28	38	37	28	31	47	100		
NC_010806-RSDaV	35	36	35	36	36	36	36	39	35	37	37	34	31	34	36	100	
NC_004750-BYDV-PAV	31	31	31	32	33	33	34	33	31	33	32	30	30	31	30	31	100

<u>Supplementary Fig. S2.</u> Pairwise aa identity of *Prunus*-infecting luteoviruses in P1-P2 (CP-RTD) fusion protein (**A**) and in P3-P5 (**B**) (viral replicase)

CHAPTER III

Molecular characterization of a new *Prunus*-infecting cheravirus and complete genome sequence of stocky prune virus.



Maryam Khalili et al. (2022)

viruses

MDPI

Article

The Molecular Characterization of a New *Prunus*-Infecting Cheravirus and Complete Genome Sequence of Stocky Prune Virus

Maryam Khalili [1], Thierry Candresse [1], Yoann Brans [2], Chantal Faure [1], Jean-Marc Audergon [3], Véronique Decroocq [1], Guillaume Roch [3] and Armelle Marais [1,*]

[1] Univ. Bordeaux, INRAE, UMR BFP, 33140 Villenave d'Ornon, France
[2] Laboratoire de Virologie et de Biologie Moléculaire, CTIFL, 24130 Prigonrieux, France
[3] UR GAFL, INRAE, 84143 Montfavet Cedex, France
* Correspondence: armelle.marais-colombel@inrae.fr; Tel.: +33-557122379

Abstract: As part of a virome characterization of *Prunus* species, a novel cheravirus was discovered in two wild species, *Prunus brigantina* and *P. mahaleb*, and in an apricot (*P. armeniaca*) accession. The sequence of the two genomic RNAs was completed for two isolates. The Pro-Pol conserved region showed 86% amino acid (aa) identity with the corresponding region of trillium govanianum cheravirus (TgCV), a tentative *Cheravirus* member, whereas the combined coat proteins (CPs) shared only 40% aa identity with TgCV CPs, well below the species demarcation threshold for the genus. This suggests that the new virus should be considered a new species for which the name alpine wild prunus virus (AWPV) is proposed. In parallel, the complete genome sequence of stocky prune virus (StPV), a poorly known cheravirus for which only partial sequences were available, was determined. A phylogenetic analysis showed that AWPV, TgCV and StPV form a distinct cluster, away from other cheraviruses.

Keywords: high-throughput sequencing; stone fruit; phylogenetic analysis; prunus; cheravirus; stocky prune virus; alpine wild prunus virus

Citation: Khalili, M.; Candresse, T.; Brans, Y.; Faure, C.; Audergon, J.-M.; Decroocq, V.; Roch, G.; Marais, A. The Molecular Characterization of a New *Prunus*-Infecting Cheravirus and Complete Genome Sequence of Stocky Prune Virus. *Viruses* **2022**, *14*, 2325. https://doi.org/10.3390/v14112325

Academic Editors: Munir Mawassi and Sergey Morozov

Received: 12 September 2022
Accepted: 19 October 2022
Published: 23 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The genus *Cheravirus* is one of the nine genera (according to the last update of https://talk.ictvonline.org/taxonomy/ verified on 3 August 2022) assigned to the family *Secoviridae*. The recognized species include *Apple latent spherical virus* (ALSV), *Cherry rasp leaf virus* (CRLV), *Arracacha virus B* (AVB), *Currant latent virus* (CuLV) and *Stocky prune virus* (StPV), while two tentative members have recently been discovered, trillium govanianum cheravirus (TgCV) and babaco cheravirus 1 (BabChV-1) [1–9]. Cheraviruses have bipartite, single-stranded positive-sense RNA genomes. Full-length recognized cheravirus RNA1 molecules have been determined in a range of 6.8–7.1 kb and RNA2 molecules in a range of 3.2–3.7 kb-long. Each genomic RNA encodes a polyprotein (RNA1: P1; RNA2: P2) that is cleaved by the RNA1-encoded 3C proteinase to generate the functional non-structural and structural proteins. Members of the genus *Cheravirus* encode three capsid proteins. Some are known to be transmitted by nematodes (CRLV) and aphids (CuLV). Seed transmissibility was observed in CRLV, ALSV and AVB, while ALSV and AVB are also pollen-transmissible [8,10–13].

Stone fruits are deciduous trees native to the temperate zone of the northern hemisphere. All stone fruits such as apricot, plum, peach/nectarine and sweet and sour cherry alongside almond belong to the genus *Prunus* of the *Rosaceae* family and can be affected by various viral diseases. CRLV and StPV are the two known *Cheravirus* members naturally infecting *Prunus*. Sweet cherry (*Prunus avium*), peach (*P. persica*) and susceptible

cherry rootstocks including both widely used "Mazzard" (*P. avium*) and "Mahaleb" (*P. mahaleb*) are the main *Prunus* hosts for CRLV [14,15]. StPV is the causal agent of the stocky prune disease that affects European plum (*P. domestica*) and that is characterized by severe symptoms of shortened internodes, chlorotic, rolled and enlarged leaves, premature fruit fall and tree decline over time [3,4,16]. The disease, geographically limited to the south-west of France, resulted in a notable loss of production. It was a source of concern at the beginning of the 20th century, but its impact has fortunately massively declined, probably as a consequence of the sanitary selection of planting materials. To date, only partial sequence data representing the 3′ part of RNA1 (NC_043387, 2644 nt) and RNA2 (NC_043388, 1794 nt) are available [4].

As part of a virus discovery effort in *Prunus* crops, the virome of over 300 samples belonging to 9 crop or ornamental species (*P. domestica*, *P. avium*, *P. cerasus*, *P. armeniaca*, *P. persica*, *P. dulcis*, *P. serrulata*, *P. nipponica*, and *P. subcordata*), as well as over 300 samples of wild *Prunus* species (*P. brigantina*, *P. mahaleb*, *P. spinosa*, *P. serotina* and *P. cerasifera*) were investigated using high-throughput sequencing (HTS). In this study, we present the molecular characterization of a new *Prunus*-infecting cheravirus identified in two wild *Prunus* species and in an apricot accession and the determination of the complete genome sequence of StPV.

2. Materials and Methods

2.1. Plant Material

Pooled leaf samples collected from different parts of the canopy, regardless of symptoms, were collected from 50 apricot (*P. armeniaca*) varieties originated from 16 countries kept in collection at INRAE *Prunus* Biological Resource Center (BRC, INRAE, Avignon, France). In particular, the apricot accession A1915 found to be infected by a novel cheravirus is an old variety (San Castrese) from the Campania region in southern Italy, introduced from Venturina Pisa University experimental repository in the collection. In addition, similar leaf samples were collected in the French Alps in 2017 [17] and 2021, respectively, for wild growing *P. brigantina* and *P. mahaleb* trees. The original, partially sequenced StPV isolate [4] was maintained in GF305 peach seedlings in collection at INRAE virus repository under greenhouse conditions. In all cases, fresh leaf tissues were stored at −80 °C or dried over anhydrous $CaCl_2$ (Sigma Aldrich Chimie, Saint-Quentin-Fallavier, France) and preserved at room temperature until use.

2.2. Double-Stranded RNA Extraction, Library Preparation and High-Throughput Sequencing (HTS)

Double-stranded RNA were purified from dried leaves of *P. brigantina* or of the 50 apricot accessions following the method of [18] before being used for cDNA synthesis using SuperscriptII Reverse Transcriptase and LDF primers [19] according to manufacturer's instructions (Invitrogen/Fisher Scientific, Illkirch, France). A random PCR amplification was performed on each cDNA preparation using multiplex identifier (MID) adaptors and Dream Taq DNA polymerase (Thermo Fisher Scientific) as described by [18], enabling all samples to be sequenced in a multiplexed format. PCR products purified using a MiniElute PCR Purification Kit (Qiagen SAS France, Courtaboeuf, France) were pooled in equimolar amounts prior to library preparation using the TruSeq Nano kit (Illumina) and Illumina sequencing on a Hiseq3000 platform (2 × 150 bp) [outsourced at the GetPlage INRAE platform (Toulouse, France) or Azenta (Leipzig, Germany)].

2.3. Total RNA (totRNA) Extraction and HTS

Total RNA HTS RNAseq was performed for a *P. mahaleb* sample and the StPV-inoculated GF305 sample. Total RNA was extracted from fresh leaves using a modified CTAB procedure [20] and sequenced after a ribodepletion step in a HiSeq3000 2 × 150 bp format (Azenta).

2.4. Total Nucleic Acids (TNAs) Extraction and RT-PCR Detection of the Novel Cheravirus

TNAs were extracted from graft-inoculated GF305 leaves as well as leaves from non-inoculated controls three months after the grafting assay (see below) using protocol 1 described in [21]. Specific primers allowing the detection of AWPV-Pa were designed from contigs in RNA1 (Seco-PA-R1-F/R, Table S1) and used to perform a two-step RT-PCR assay as already described [22].

2.5. HTS Data Analyses

Sequencing reads were quality controlled and trimmed before de novo assembly using CLC Genomic Workbench version 21.0.3 (Qiagen). A BlastX analysis to identify viral contigs was performed against the GenBank protein database (nr) limited to viruses. The contigs detected as representing new viral sequences were extended by rounds of mapping residual reads. If needed, extended contigs were manually scaffolded before closing any residual small gaps by RT-PCR and the Sanger sequencing of amplicons. An additional analysis was conducted by mapping HTS reads on a set of *Prunus* viruses reference genomes (min length fraction = 0.8; min similarity fraction = 0.7).

2.6. Genome Sequence Completion for Isolates of the Novel Cheravirus and StPV

Based on the scaffold sequence reconstructed for the RNA1 of the novel cheravirus *P. mahaleb* isolate, a primer pair (Table S1) was designed to amplify a cDNA fragment spanning the remaining small gap in the sequence assembled for the *P. brigantina* isolate, using a two-step RT-PCR assay as described in [22]. The 5′ ends of the two genomic RNAs of StPV and of the *P. brigantina* and *P. mahaleb* isolates of the novel cheravirus were obtained by the rapid amplification of cDNA ends (RACE) experiments using the SMARTer® RACE 5′/3′Kit (Takara Bio Europe SAS, Saint-Germain-en-Laye, France) on either dsRNA or totRNA templates using primers designed based on the assembled HTS sequences (Table S1) and following the manufacturer's instructions. Additionally, the 3′ ends were amplified using contig-based designed forward primers together with LD-prim primers (Table S1) through long-distance (LD) RT-PCR according to the protocol already described [23]. Amplicons were directly Sanger sequenced on both strands (Eurofins Genomics, Ebersberg, Germany).

2.7. Pairwise Comparisons, Phylogenetic and Recombination Analyses

Multiple alignments of nucleotide (nt) or amino acid (aa) sequences were performed using the ClustalW program [24] in MEGA 7 [25]. The aa alignments were used for maximum likelihood phylogenetic analyses using the RtRev +I+G evolutionary model. Branch validity was evaluated by randomized bootstrapping (1,000 replicates). Conserved protein domains in the aa deduced sequences were searched using the CD search tool (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi, accessed on 22 August 2022). Recombination analysis was performed using alignments of RNA1 or RNA2 of cheraviruses and the RDP4 program with default parameters [26].

2.8. Biological Indexing in Greenhouse

Wood samples from the apricot accession A1915 were sampled to conduct a biological indexing in greenhouse. Part of the tree showed leaf deformation and chlorosis symptoms. Biological indexing on peach GF305 with chip-budding graft inoculation was conducted [27]. Two weeks after grafting, the GF305 indicators were cut off, and ten replicates (five from symptomatic and five from non symptomatic branches) were monitored over several weeks to spot the appearance of any virus-like symptoms on the growth and leaves of the indicator plants in comparison to negative controls (non inoculated GF305).

3. Results

3.1. Determination of the Complete Genome Sequence of a Newly Discovered Cheravirus and of StPV

Of several *P. brigantina* samples analyzed by dsRNA-based HTS, one revealed the presence of a novel *Secoviridae*. After demultiplexing, quality trimming and the de novo assembly of the 18,93,230 reads obtained for this sample, BlastX-based contigs annotation allowed us to identify contigs with significant identity with members of the *Secoviridae* family. Scaffolds for both genomic RNAs were then reconstructed using the contigs of interest. Consecutive rounds of reads mapping to the scaffolds extended the sequences to yield near-complete viral molecules. A small gap in RNA1 scaffold was filled by the direct sequencing of a RT-PCR fragment generated from dsRNA using a primer pair flanking the gap (Table S1). Overall, the reads assembled in the completed sequences represented 30.5% and 22% of total reads for RNA1 and RNA2, respectively, corresponding to average coverages of 8,886× (RNA1) and 15,047× (RNA2) (Table 1). In Blast searches of the GenBank database, the most closely related virus was trillium govanianum cheravirus (TgCV), a tentative member in the *Cheravirus* genus with 54% nt identity in RNA1 and 56.2% nt identity in RNA2. When used as references, the scaffolds assembled for the *P. brigantina* virus enabled us to identify two other related viruses in two HTS datasets: one from a *P. mahaleb* tree from the French Alps, one for which two long contigs representing nearly complete genomic sequences (RNA1 and 2) were obtained at high coverage (Table 1), and the other from an apricot (*P. armeniaca*) sample from Prunus BRC for which the assembly was much less complete. For the RNA1 of the isolate in the apricot sample, 19 small nonoverlapping contigs were obtained, covering an overall 4650 nt of the RNA1 molecule. The RNA2 molecule was even more fragmented. Further analyses revealed that the two isolates identified in *P. brigantina* and *P. mahaleb* share 86.3% nt identity in their RNA1 (93.7% aa identity in the encoded polyprotein) and 86.7% nt identity in their RNA2 (93.6% aa identity in the polyprotein). More precisely, in the two regions used for molecular species demarcation in the *Secoviridae* family (Pro-Pol region and combined CPs), the levels of identity between the two isolates are well above the cut-off (less than 80% or 75% aa identity, respectively, [28]), with respectively 97.1% and 93.9% aa identity (Table 2). In parallel, similar pairwise comparisons performed with the assembled scaffold of the RNA1 of the isolate identified in apricot comprising 4650 nt showed that it shares 84.3–85% nt identity with the corresponding region of the RNA1 of the *P. mahaleb* and *P. brigantina* isolates (93.6–93.2% for the encoded polyprotein fragment). These findings show that the three identified viruses are isolates of a novel *Cheravirus* species that naturally infects different *Prunus* hosts and tentatively named alpine wild prunus virus (AWPV). RACE experiments were performed to determine the 5′ and 3′ end sequences for both genomic RNAs of the *P. mahaleb* and *P. brigantina* isolates, while no specific effort was made to complete the genome sequence of the apricot isolate. The full-length (excluding polyA tail) RNA1 of AWPV in *P. brigantina* and *P. mahaleb* was determined to be 7354 nt and 7491 nt, respectively and the RNA2 was 3574 nt and 3568 nt long, respectively (Table 3). The complete genome sequences have been deposited in GenBank under the accession numbers OP328247-8 for AWPV from *P. brigantina* (AWPV-Pb), and OP328249-50 for AWPV from *P. mahaleb* (AWPV-Pm) (Table 1).

Table 1. Target nucleic acids population, number of trimmed reads, average coverage and percentage of total reads mapping to the genomic RNAs of alpine wild prunus virus (AWPV) and stocky prune virus (StPV).

Virus-Isolate [1]	Method	Trimmed Reads	Segment	Average Coverage	Mapped Reads (%)	Accession Numbers
AWPV-Pb	dsRNA [2]	1,893,230	RNA1	8886x	30.5%	OP328247
			RNA2	15047x	22%	OP328248
AWPV-Pm	RNA	61,886,192	RNA1	3503x	0.2%	OP328249
			RNA2	10145x	0.39%	OP328250
StPV	RNA	57,041,248	RNA1	4051x	0.34%	OP328251
			RNA2	4476x	0.18%	OP328252

[1] AWPV-Pb: alpine wild prunus virus-*P. brigantina* ; AWPV-Pm: alpine wild prunus virus-*P. mahaleb* ; StPV: stocky prune virus. [2] dsRNA: double-stranded RNA.

Table 2. Percentages of identity observed between the Pro-Pol region and the capsid protein region of alpine wild prunus virus-*P. brigantina* (AWPV-Pb) and the corresponding proteins of *Cheravirus* accepted or tentative members.

Virus [1]	Amino Acid Identity (%)	
	Pro-Pol [2]	Capsid Protein [3]
AWPV-Pm	97.1%	93.9%
StPV	64.7%	33.7%
ALSV	38.7%	16.3%
AVB	41.6%	12.2%
CRLV	38.9%	15.9%
CuLV	39.4%	14.6%
TgCV	86.3%	41.1%
BabChV-1	39.2%	na [4]

[1] AVB: arracacha virus B; ALSV: apple latent spherical virus; BabChV-1: babaco cheravirus-1; CRLV: cherry rasp leaf virus; CuLV: currant latent virus; TgCV: trillium govanianum cheravirus. [2] The Pro-Pol region is delineated by the "CG" motif of the 3C-like proteinase and the "GDD" motif of the polymerase [28]. [3] For viruses with two or three CPs, combined CP sequences are considered [28]. [4] not applicable (No RNA2 data available in GenBank).

Table 3. Characteristics of the genomic RNAs of the cheraviruses characterized in this study.

Virus	Segment	Length (nt)	Polyprotein (aa)	5′ NCR [1] (nt)	3′ NCR (nt)
AWPV-Pb	RNA1	7354	2358	50	230
	RNA2	3574	1088	69	240
AWPV-Pm	RNA1	7491	2404	50	229
	RNA2	3568	1086	70	241
StPV	RNA1	7021	2230	80	251
	RNA2	3495	1056	78	249

[1] NCR: noncoding region.

In parallel, a total of 57,041,248 paired-end Illumina reads were recovered for the StPV graft-inoculated GF305 sample (Table 1). Following the de novo assembly of trimmed reads and BlastX analyses, two long contigs of nearly complete genome length were identified for the genomic RNAs of StPV, with over 4000x coverage (Table 1). Both sequences showed 99.9% nt identity with the previously determined partial StPV RNA1 and RNA2 sequences (NC_043387-88). The 5′ and 3′ ends were determined by RACE or polyA-anchored PCR for each molecule. The complete StPV RNA1 (7021 nt-long) and RNA2 (3495 nt-long) sequences have been deposited in GenBank under accession numbers OP328251 and OP328252, respectively (Table 1).

3.2. Genomic Organization of AWPV and StPV and Determination of Their Phylogenetic Relationships

Based on the known genomic organization of other cheraviruses and on sequence homologies with them, a genomic organization of AWPV and StPV was proposed (Figure 1A,B) in which polyprotein 1 (P1) is cleaved in six mature proteins, while polyprotein 2 (P2) is cleaved into four proteins including a movement protein and three coat proteins. However, the analysis of polyprotein alignments and the variety of cleavage sites already reported for cheraviruses [1,2,6–8] did not allow for predicting cleavage sites for StPV and AWPV polyproteins. A single open reading frame (ORF) is identified in both RNA1 and RNA2 (Figure 1A,B). The AUG start codon at nt positions 51–53 in RNA1 AWPV-Pb and AWPV-Pm (nt 81–83 in RNA1 StPV) is in favorable translation initiation context, with a purine (A/G) at position -3, [29] for both AWPV and StPV. In the case of RNA2, although the AUG codon in nt positions 70–72, 71–73 and 79–81, in AWPV-Pb, AWPV-Pm and StPV, respectively, is in a suboptimal context for initiating translation, it is considered to be the most likely site of initiation, since 5′ NCRs would then being the same length range in both RNA1 and RNA2, while the use of the next downstream AUG codon (286–288, 287–289 and 255–257, respectively in AWPV-Pb, AWPV-Pm and StPV) would result in 5′ NCRs widely differing in length between the two genomic RNAs. The RNA1 and RNA2 segments of AWPV-Pb and AWPV-Pm are very largely colinear, but in the RNA1 sequence, an indel of 138 nt is located between nt 278 and 416 (referring to the AWPV-Pm RNA1), leading to a deletion of 46 aa in the N-terminal part of deduced P1 polyprotein in AWPV-Pb. The RNA2 also presents a short indel of six nt in the *P. mahaleb* isolate (from nt 2077 to 2083, referring to AWPV-Pb RNA2), leading to a deletion of two aa in the deduced P2 polyprotein in the CPs coding region.

The RNA1 5′ noncoding regions (NCRs) are 50 nt long in AWPV-Pb and AWPV-Pm, and 80 nt long in StPV (Table 3). The 3′ NCRs are longer, comprising between 229 nt (AWPV-Pm) and 251 nt (StPV) (Table 3). In RNA2, the 5′ NCRs are 69, 70 and 78 nt long, and the 3′ NCRs comprise between 241 nt (AWPV-Pb) and 251 nt (AWPV-Pm) (Table 3). As reported for other *Secoviridae* members with bipartite genomes, sequence homologies were identified between the NCRs of the two genomic RNAs. The 5′ NCRs show 22 to 28 fully conserved 5′ nucleotides, depending on the virus considered, whereas the 3′ NCRs display 89.5%, 97.8% and 98.9% sequence identity between the two genomic RNAs of StPV, AWPV-Pb, and AWPV-Pm, respectively.

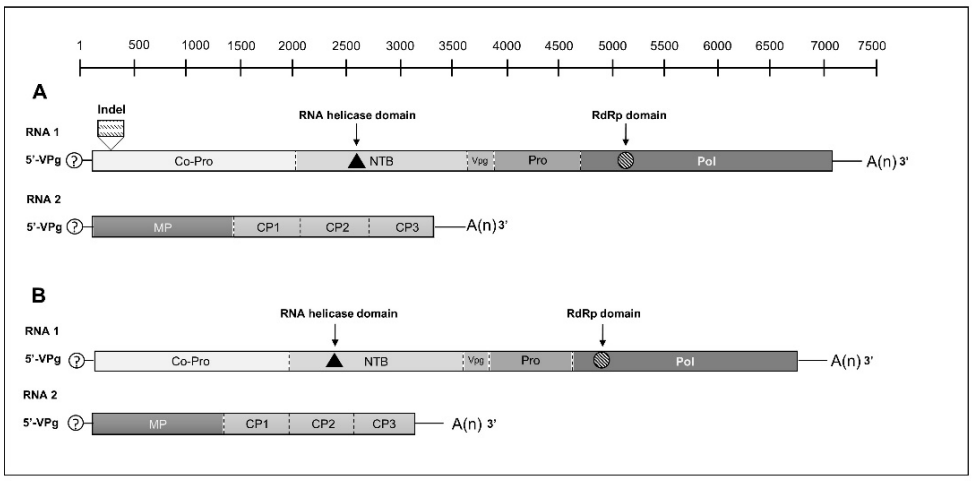


Figure 1. Hypothetical genome organization of alpine wild prunus virus-*P. brigantina* (AWPV-Pb) (**A**) and of stocky prune virus (StPV) (**B**). Putative cleavage sites of RNA1 and 2 deduced polyproteins are represented by dashed lines. The conserved motifs of helicase and RNA-dependent RNA polymerase (RdRp) are shown by black triangle and hatched circle, respectively. The position of the indel of 138 nt in AWPV-Pb RNA1 in comparison to AWPV-Pm RNA1 is shown. Co-Pro: protease cofactor; NTB: nucleotide triphosphate binding helicase; VPg: viral genome-linked protein; Pro: protease; Pol: polymerase; MP: movement protein; CP: coat protein. RdRp: RNA-dependent RNA polymerase.

As expected, two conserved domains were identified within the P1 polyprotein of both AWPV and StPV. The first was an RNA helicase domain (pfam 00910) located between aa 849 and 949 in AWPV-Pb. Corresponding values for AWPV-Pm and StPV are, respectively, 895–995 and 835–933. The second was an RNA-dependent RNA polymerase domain (cd01699) between aa 1731–2027 for AWPV-Pb and aa 1777–2073 or 1712–2011 for AWPV-Pm and StPV, respectively. No conserved domains could be identified in the P2 polyproteins, which is in keeping with the very low percentages of identity observed between the CPs of different cheraviruses (Table 2).

An analysis of potential recombination events in the genomes of *Cheravirus* members or potential members using RDP4 [26] failed to identify any significant events. In order to confirm the taxonomical position of AWPV, phylogenetic analyses based on the P1 and P2 polyproteins of representative *Cheravirus* members were performed (Figure 2). The phylogenetic trees clearly show the clustering of the two isolates of AWPV together with StPV and TgCV, within a subgroup distinct from the rest of the cheraviruses (Figure 2). The same clustering was obtained when a phylogenetic tree integrating representative members of the various genera in the *Secoviridae* family was generated based on the Pro-Pol region (Figure S1). Pairwise sequence comparisons showed AWPV-Pb Pro-Pol aa sequence shares between 38.7% (with ALSV) and 86.3% (with TgCV) aa identity (Table 2). Separately, the level of identity between CPs is lower, comprising between 12.2% (with AVB) and 41.1% (with TgCV) (Table 2). The pairwise comparisons performed with the AWPV-Pm provided similar values of identity. Concerning StPV, the most closely related cheraviruses are AWPV (Pb and Pm) and TgCV with 64–64.7% and 61.7% aa identity in the Pro-Pol region, respectively, and 33.8–34.1% and 40% aa identity in the CPs region, respectively.

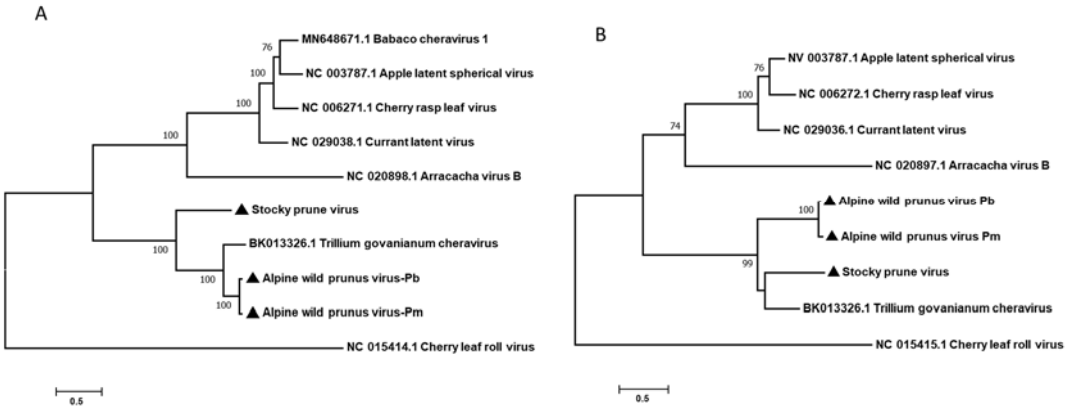


Figure 2. Phylogenetic trees reconstructed using the alignment of amino acid sequences of RNA1-encoded polyprotein P1 (**A**), and RNA2-encoded polyprotein P2 (**B**) of representative members of the genus *Cheravirus*. Trees were generated using the Maximum Likelihood method. The statistical significance of branches was evaluated by bootstrap analysis (1000 replicates). Cherry leaf roll virus (CLRV) was used as an outgroup. Bootstrap values less than 70% were removed. Viruses characterized in the present study are marked by black triangles. AWPV-Pb: alpine wild prunus virus-*P. brigantina*; AWPV-Pm: alpine wild prunus virus prunus-*P. mahaleb*; StPV: stocky prune virus; AVB: arracacha virus B; ALSV: apple latent spherical virus; BabChV-1: babaco cheravirus-1; CRLV: cherry

rasp leaf virus; CuLV: currant latent virus; TgCV: trillium govanianum cheravirus. Accession numbers are given before virus abbreviations. The bars represent the genetic distances.

3.3. Biological Indexing of AWPV-Infected Apricot Accession

About three months after biological indexing in greenhouse, some GF305 indicator plants showed symptoms of leaf chlorosis, strangulation and deformation. The presence of AWPV-Pa was verified by RT-PCR. All the GF305 plants, with and without symptoms, were found to be infected by AWPV-Pa. The observed symptoms may be associated with the presence of prunus necrotic ringspot virus (PNRSV), which was confirmed by DAS-ELISA testing. These symptoms could therefore not be linked to the presence of AWPV.

4. Discussion

Metagenomics have greatly expanded our knowledge of plant viromes by contributing to the discovery of novel viruses. This includes the viromes shared in both wild and cultivated crops. Taking advantage of HTS, we investigated a large number of wild and cultivated *Prunus* samples, leading to the discovery of a putative new cheravirus tentatively named alpine wild prunus virus (AWPV), in three different host species, *P. brigantina* (wild relative of apricot), *P. armeniaca* (cultivated apricot) and *P. mahaleb* (mahaleb cherry or St. Lucie cherry). Furthermore, this paper reports the complete genome sequence of two isolates of AWPV and the completion of the genome sequence of the previously partially characterized StPV [4]. These findings add AWPV to the list of tentative cheraviruses, which already increased over the last few years with CuLV [8], BabChV-1 [5] and TgCV [9].

The size of the AWPV RNA1 is slightly above the expected length (6.8–7.1 kb) of RNA1 in the genus *Cheravirus*. This can be explained by considering that only four complete genomes of cheraviruses had been available when establishing the genome content [28]. Although the percentage of the aa identity between AWPV and TgCV in the Pro-Pol region is above the species demarcation threshold currently accepted in the *Secoviridae* family (86.3% vs. 80%), the level of aa identity in the CPs region is well below the cut-off (41.1% vs. 75%), so that there is no ambiguity that AWPV represents a distinct species.

AWPV was initially identified in a *P. brigantina* tree sampled in 2017 in a rather isolated area of the French Alps. Efforts to resample that tree in 2021 were not successful since the tree had died in the interim period from unknown causes. Sampling in 2021 of *P. mahaleb* trees in a nearby valley, about 25 km away on a direct line, allowed us to identify one tree out of the 10 sampled which was infected by AWPV. AWPV was also later identified in a single apricot tree of the old Italian variety San Castrese. Even though the virus was detected in single infection in *P. brigantina* and in co-infection with another virus (prunus mahaleb-associated luteovirus, PmaLV) in *P. mahaleb* [30], the association of the symptoms with AWPV infection was not possible because of the unusual harsh, high mountain, growth conditions of the trees. Both the dead *P. brigantina* tree and the *P. mahaleb* tree showed a bushy growth and shortened internodes, but it is not possible to know whether these symptoms were caused by AWPV infection or by mountain growth conditions, or in the case of the *P. mahaleb* tree, by the co-infection with PmaLV, a very recent discovered luteovirus [30]. It should be noted, though, that of the 10 *P. mahaleb* trees sampled in 2021, which all showed similar poor, stunted growth, a single one was found infected by AWPV. The leaf deformation and chlorosis observed on the leaves of the AWPV infected apricot could likewise be correlated with its co-infection by PNRSV. The results from the bioassays conducted on samples collected from this apricot tree came to the same conclusion. Further studies would be necessary, especially from AWPV single-infected plant material, to assess the symptomatology associated with this virus. Overall, our observation cannot be used to establish a clear link between AWPV infection and symptoms in *Prunus* hosts.

Unlike AWPV, StPV is known to be the causal agent of a severe disease developing leaf chlorosis and deformation, bushy growth and early drop of the fruits, symptoms

diversely called "maladie de Brugères", "maladie du prunier stérile", "maladie du prunier mâle" or stocky prune disease [3,4,16]. It also induces particularly severe stunting in the GF305 peach indicator, and there is overall no ambiguity about its pathogenicity in *Prunus*. The completion of its genomic sequence reported here unambiguously confirms that it belongs to the *Cheravirus* genus and demonstrates that it is related to AWPV.

The availability of full genomic sequences facilitates the development of specific and inclusive detection tests, which would permit further studies on the geographical distribution of AWPV, its genetic diversity and its host range. The fact that it was identified in two different *Prunus* species sampled each time as few individuals suggests that it might not be rare in wild *Prunus* hosts in that particular area of the French Alps and possibly elsewhere. This virus was detected in an area with few or no *Prunus* crops, which may have prevented its transfer from wild to cultivated *Prunus* species. Ultimately, this could be one of the reasons it was detected only once among more than 300 cultivated *Prunus* accessions tested by HTS during this survey. In a similar fashion, StPV was not detected among the more than 600 *Prunus* samples analyzed by HTS, which is in keeping with its initial limited distribution and (near) complete disappearance in recent times. Although soil transmission was considered a possibility for StPV, no vector has been identified to date and the transmission mechanism(s) of AWPV remain to be analyzed.

Once again, this study demonstrates the power of HTS approaches for virus discovery. A precise evaluation of the risks potentially associated with AWPV for *Prunus* crops remains to be performed, in particular when it comes to pathogenicity to a range of cultivated *Prunus* species and to the existence and efficiency of potential transmission mechanisms.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/v14112325/s1, Figure S1: Phylogenetic tree reconstructed using the alignment of amino acid sequences of the Pro-Pol region (conserved domain between the "CG" motif of the 3C-protease and the "GDD" motif of the RdRp) of representative members of the family *Secoviridae*. The tree was generated using the neighbor joining method and strict aa identity p-distances. The statistical significance of branches was evaluated by bootstrap analysis (1000 replicates). Bootstrap values less than 70% were removed. The scale bar represents 5% amino acid divergence. Accession numbers and names of the viruses included in the tree are indicated. Genera to which viruses belong are indicated at the right of the figure. Table S1: List of primers used in this study, their purpose and their target molecule.

Author Contributions: Conceptualization, M.K., T.C., A.M. and Y.B.; methodology, M.K., C.F.; resources, J.-M.A., G.R., V.D., T.C. and Y.B.; writing—original draft preparation, M.K. and A.M.; writing—review and editing, M.K., A.M. and T.C. All authors have read and agreed to the published version of the manuscript.

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). The plant indexing biological tests conducted by CTIFL were funded by INTERFEL (Fresh fruit and vegetable interprofessional association).

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable

Data Availability Statement: The sequences reported in the present manuscript have been deposited in the GenBank database under the accession numbers OP328247-OP328252.

Acknowledgments: The authors would like to thank the platform GetPlaGe (GenoToul, INRAE, Toulouse, France) for the Illumina sequencing, and P. Briard (UMR BFP, INRAE, Bordeaux, France) and F. Latour (CTIFL) for plant indexing.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Adams, I.P.; Glover, R.; Souza-Richards, R.; Bennett, S.; Hany, U.; Boonham, N. Complete genome sequence of arracacha virus B: A novel cheravirus. *Arch. Virol.* **2013**, *158*, 909–913. https://doi.org/10.1007/s00705-012-1546-x.
2. Thompson, J.R.; Perry, K.L.; De Jong, W. A new potato virus in a new lineage of picorna-like viruses. *Arch. Virol.* **2004**, *149*, 2141–2154. https://doi.org/10.1007/s00705-004-0362-3.
3. Candresse, T.; Delbos, R.; Le Gall, O.; Dunez, J.; Desvignes, J. Characterization of stocky prune virus, a new nepovirus detected in french prunes. *Acta Hortic.* **1998**, *472*, 175–182. https://doi.org/10.17660/ActaHortic.1998.472.18.
4. Candresse, T.; Svanella-Dumas, L.; Le Gall, O. Characterization and partial genome sequence of stocky prune virus, a new member of the genus *Cheravirus*. *Arch. Virol*. **2006**, *151*, 1179–1188. https://doi.org/10.1007/s00705-005-0682-y.
5. Cornejo-Franco, J.F.; Medina-Salguero, A.; Flores, F.; Chica, E.; Grinstead, S.; Mollov, D.; Quito-Avila, D.F. Exploring the virome of Vasconcellea x heilbornii: The first step towards a sustainable production program for babaco in Ecuador. *European J. Plant Pathol.* **2020**, *157*, 961–968. https://doi.org/10.1007/s10658-020-02037-2.
6. James, D.; Upton, C. Nucleotide sequence analysis of RNA-2 of a flat apple isolate of Cherry rasp leaf virus with regions showing greater identity to animal picornaviruses than to related plant viruses. *Arch. Virol.* **2002**, *147*, 1631–1641. https://doi.org/10.1007/s00705-002-0833-3.
7. Li, C.; Yoshikawa, N.; Takahashi, T.; Ito, T.; Yoshida, K.; Koganezawa, H. Nucleotide sequence and genome organization of Apple latent spherical virus: A new virus classified into the family *Comoviridae*. *J. Gen. Virol*. **2000**, *81*, 541–547. https://doi.org/10.1099/0022-1317-81-2-541.
8. Petrzik, K.; Přibylová, J.; Špak, J.; Havelka, J. Partial genome sequence of currant latent virus, a new chera-like virus related to Apple latent spherical virus. *J. Gen. Plant Pathol.* **2015**, *81*, 142–145. doi.org/10.1007/s10327-014-0574-7.
9. Sidharthan, V.K.; Kalaivanan, N.S.; Baranwal, V.K. Discovery of putative novel viruses in the transcriptomes of endangered plant species native to India and China. *Gene* **2021**, *786*, 145626. https://doi.org/10.1016/j.gene.2021.145626.
10. Hansen, A.J.; Nyland, G.; McElroy, F.D.; Stace-Smith, R. Cherry rasp leaf disease in North America. *Phytopathology* **1974**, *64*, 721–727.
11. Jones, R.A.C..Tests for transmission of four potato viruses through potato true seed. *Ann. Appl. Biol*. **1982**, *100*, 315–320. doi.org/10.1111/j.1744-7348.1982.tb01944.x.
12. Nakamura, K.; Yamagishi, N.; Isogai, M.; Komori, S.; Ito, T.; Yoshikawa, N. Seed and pollen transmission of Apple latent spherical virus in apple. *J. Gen. Plant Pathol*. **2011**, *77*, 48–53. https://doi.org/10.1007/s10327-010-0275-9.
13. Nyland, G.; Lownsbery, B.F.; Lowe, S.K.; Mitchell, J.F. The transmission of cherry rasp leaf virus by *Xiphinema americanum*. *Phytopathology* **1969**, *59*, 1111–1112.
14. Stace-Smith, R.; Hansen, A.J. Some properties of cherry rasp leaf virus. *Acta Hort.* **1976**, *67*, 193–198. https://doi.org/10.17660/ActaHortic.1976.67.23.
15. Wagnon, H.K. Investigations of cherry rasp leaf disease in California. *Plant Dis. Rep.* **1968**, *52*, 618–622.
16. Desvignes, J.C. Stocky prune. In *Les Virus des Arbres Fruitiers*; Desvignes, J.C. Ed.; CTIFL: Paris, France, 1990; pp. 51–52.
17. Liu, S.; Decroocq, S.; Harte, E.; Tricon, D.; Chague, A.; Balakisiyeva, G.; Kostriesyna, T.; Turdiev, T.; Fisher-Le Saux, M.; Dallot, S.; et al. Genetic diversity and population structure analyses in the Alpine plum (Prunus brigantina Vill.) confirm its affiliation to the Armeniaca section. *TGG* **2021**, *17*, 2. https://doi.org/10.1007/s11295-020-01484-6.
18. Marais, A.; Faure, C.; Bergey, B.; Candresse, T. Viral Double-Stranded RNAs (dsRNAs) from Plants: Alternative Nucleic Acid Substrates for High-Throughput Sequencing. *Methods Mol Biol.* **2018**, *1746*, 45–53. https://doi.org/10.1007/978-1-4939-7683-6_4.
19. François, S.; Filloux, D.; Fernandez, E.; Ogliastro, M.; Roumagnac, P. Viral Metagenomics Approaches for High-Resolution Screening of Multiplexed Arthropod and Plant Viral Communities. *Methods Mol Biol.* **2018**, *1746*, 77–95. https://doi.org/10.1007/978-1-4939-7683-6_7.
20. Chang, S.; Puryear, J.; Cairney, J. A simple and efficient method for isolating RNA from pine trees. *Plant Mol. Biol. Rep.* **1993**, *11*, 113–116. https://doi.org/10.1007/BF02670468.
21. Foissac, X.; Svanella-Dumas, L.; Gentit, P.; Dulucq, M.J.; Marais, A.; Candresse, T. Polyvalent degenerate oligonucleotides reverse transcription-polymerase chain reaction: A polyvalent detection and characterization tool for trichoviruses, capilloviruses, and foveaviruses. *Phytopathology* **2005**, *95*, 617–625.
22. Marais, A.; Faure, C.; Couture, C.; Bergey, B.; Gentit, P.; Candresse, T. Characterization by deep sequencing of divergent Plum bark necrosis stem pitting-associated virus (PBNSPaV) isolates and development of a broad-spectrum PBNSPaV detection assay. *Phytopathology* **2014**, *104*, 660–666. https://doi.org/10.1094/PHYTO-08-13-0229-R.
23. Youssef, F.; Marais, A.; Faure, C.; Barone, M.; Gentit, P.; Candresse, T. Characterization of Prunus-infecting Apricot latent virus-like Foveaviruses: Evolutionary and taxonomic implications. *Virus Res.* **2011**, *155*, 440–445. https://doi.org/10.1016/j.virusres.2010.11.013.
24. Thompson, J.D.; Higgins, D.G.; Gibson, T.J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673–4680. https://doi.org/10.1093/nar/22.22.4673.
25. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol*. **2016**, *33*, 1870–1874. https://doi.org/10.1093/molbev/msw054.

26. Martin, D.P.; Murrell, B.; Golden, M.; Khoosal, A.; Muhire, B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* **2015**, *1*, vev003. https://doi.org/10.1093/ve/vev003.
27. Gentit, P.; Foissac, X.; Svanella-Dumas, L.; Peypelut, M.; Macquaire, G.; Candresse, T. Biological properties and partial characterisation of two different Foveavirus associated with similar disorders of cherry trees. *Acta Hortic* **2001**, *550*, 161–168. https://doi.org/10.17660/ActaHortic.2001.550.23.
28. Thompson, J.R.; Dasgupta, I.; Fuchs, M.; Iwanami, T.; Karasev, A.V.; Petrzik, K.; Sanfaçon, H.; Tzanetakis, I.; van der Vlugt, R.; Wetzel, T.; et al. ICTV Virus Taxonomy Profile: Secoviridae. *J. Gen. Virol.* **2017**, *98*, 529–531. https://doi.org/10.1099/jgv.0.000779.
29. Kozak, M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **1986**, *44*, 283–292. https://doi.org/10.1016/0092-8674(86)90762-2.
30. Khalili, M.; Candresse, T.; Koloniuk, I.; Safarova, D.; Brans, Y.; Faure, C.; Delmas, M.; Massart, S.; Aranda, M.A., Cagalayan, K.; et al. The expanding menagerie of *Prunus*-infecting luteoviruses. *Phytopathology* **2022**. https://doi.org/10.1094/PHYTO-06-22-0203-R.

CHAPTER IV

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

Maryam Khalili et al. (2022)

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

M. Khalili[1], T. Candresse[1], C. Faure[1], Y. Brans[2], A. Marais[1*],

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

[2] *Laboratoire de Virologie et de Biologie moléculaire, CTIFL, 24130 Prigonrieux, France*

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

ORCID ID: 0000-003-2482-1543

Abstract: 186 words

Text (incl. references & figures legends): 2,098 words

Figures and Tables: 2 figures and 2 tables

The nucleotide sequences reported in this manuscript have been deposited in GenBank under accession numbers XXXX

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

Abstract

We report here the complete genome sequence of two biologically characterized *Prunus*-infecting nepoviruses, apricot latent ringspot virus (ALRSV) and myrobolan latent ringspot virus (MLRSV), for which no or very little genomic information was available. Phylogenetic analyses confirm they both belong to the subgroup C of the genus *Nepovirus*. Pairwise comparisons show they share only 57% of amino acid (aa) identity in their coat protein, which fits with absence of cross reaction in serology. However, due to the similar symptoms in woody and herbaceous hosts they cause, as well as the level of aa identity they share in the Pro-Pol region, the question to know whether they represent distinct species remains. In addition, as part of a small-scale datamining effort to screen wild and botanical *Prunus* public RNASeq data, two contigs representing the nearly complete genome of a new nepovirus have been assembled from a dataset (SRR8369794) from the smooth stone peach *Prunus mira* in the Himalayas. Phylogenetic analyses, pairwise comparisons with *Nepovirus* members and predicted genome organization suggested this novel species, tentatively named prunus mira virus A, belongs to the subgroup A of the genus *Nepovirus*.

Members of the *Nepovirus* genus in the *Secoviridae* family (subfamily *Comovirinae*) have bipartite single-stranded, positive-sense RNA genomes encapsidated in icosaedral particles involving a single, large coat protein (CP) subunit. They are divided into three subgroups, A, B, and C based on their genome size and organization, and on their polyprotein's cleavage sites [1]. Myrobolan latent ringspot virus (MLRSV) was described from a *Prunus cerasifera* tree (Myrobolan) showing very poor growth and reduced vigor [2]. While having a known geographical distribution limited to a small area of southwest France, it 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 nepoviruses, characterized by a large RNA2 molecule now composing the genus subgroup C [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 orchard showing reduced foliage and a bare, skeletal 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 new subgroup C *Nepovirus*. Whereas significant molecular and biological data is available for these two viruses [2-6], only a very partial RNA2 genomic sequence [6] is available for ALRSV (2,218 nt, AJ278875), while no genomic sequence information is available for MLRSV.

ALRSV and MLRSV isolates were maintained by grafting of GF305 peach seedling indicators and leaves from infected plants used for total RNAs extraction [7]. RNAs were converted to cDNA, ribodepleted and sequenced (HiSeq3000, 2x150nt paired reads; *ca.* 30 million and *ca*. 49 million reads, respectively). Following a trimming on quality, reads were assembled using CLC Genomics

Workbench 21.0.3 (CLC GW, Qiagen), and contigs annotated by Blastn and Blastx against GenBank. In parallel, a small-scale datamining effort was performed as described above on publicly available RNASeq Sequencing Read Archives (SRAs) from wild or botanical *Prunus* species. One SRA (SRR8369794, *ca.* $48x10^6$ paired reads) derived from smooth stone peach (*Prunus mira*) from western Himalayas was identified as yielding two long nepovirus-like contigs (BlastX e-value: 0.0). Identified contigs were validated and extended by repeated rounds of mapping of residual reads using CLC GW.

Two large contigs representing the two genomic RNAs of ALRSV and MLRSV were identified, missing only short terminal sequences (Table 1). 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 (AN XXX) and 7,678 nt (AN XXX), excluding poly A tail, for ALRSV and MLRSV respectively, while their RNA2 were respectively 5,686 nt (AN XXX) and 5,736 nt (AN XXX). Similarly, two long contigs representing near a complete genome (RNA1: 7,851 nt, AN XXX; RNA2: 3,814 nt, AN XXX) were reconstructed from the SRR8369794 *P. mira* SRA data, showing closest affinities to the nepovirus raspberry ringspot virus (RpRSV) with 64% and 63% amino acid (aa) identity in the deduced P1 and P2 polyproteins, respectively. These identity levels are clearly below the currently accepted molecular species cut-off values in the *Nepovirus* genus, suggesting that these contigs could represent the genome of 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. On the other hand, the size of the 3' UTRs of both molecules in ALRSV and MLRSV are quite similar (1,233

nt and 1,257 nt in RNA1 and RNA2 ALRSV, and 1,236 nt and 1,278 nt in RNA1 and RNA2 MLRSV), whereas significantly shorter in PMVA (597 nt and 320 nt in RNA1 and RNA2, respectively) The nucleotide sequences of the 3' ends (320 nt) of each RNA segment are nearly identical (97%) in PMVA. The 3' UTRs of MLRSV RNA1 and RNA2 also share 97%, while the 3' UTRs of ALRSV RNA1 and RNA2 are 100% identical.

Analysis of recombination events performed using alignments of RNA1, or RNA2 of nepoviruses in RDP4 program [8] with default parameters failed to detect any significant events RNA1 and RNA2. In addition, a pairwise comparison was performed to determine whether a reassortment process drives the recombination event that may be leading to the emergence of ALRSV and MLRSV. The analysis demonstrated that RNA1 and RNA2 in ALRSV and MLRSV share 80% and 74% identity, respectively (Table 2).

The three genomes have a typical nepovirus organization [1] with a unique long open reading frame (ORF) encoded on each RNA (Fig. 1). The putative P1 polyproteins of 2,138, 2,136 and 2,376 aa for respectively ALRSV, MLRSV and PMVA harbor as expected two conserved domains: a helicase domain (pfam00910) and RNA-dependent RNA polymerase domain (RdRp, cd01699) as shown in Fig. 1. The putative polyproteins P2 are 1,452, 1,459 and 1,107 aa-long, respectively and contain one to three conserved coat protein (CP) domains, depending on the virus (Fig. 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] (Fig. 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 polyprotein P1 could be confidently predicted, while for the two sites towards the 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 (Fig. 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 closely related to RpRSV (Fig. 2). A similar phylogenetic analysis performed using the Pro-Pol region provided the same clustering of ALRSV, MLRSV and PMVA (data not shown), while ALRSV and MLRSV much more closely related. Indeed, ALRSV and MLRSV share 57% aa identity in their CP (below the 75% species cut-off) but 94% aa identity in their P1 polyprotein Pro-Pol region (above the 80% species cut-off) [1]. Similar analyses performed with PMVA and related subgroup A *Nepovirus* members showed aa identity ranges of 38-70% in the Pro-Pol region and 19-57% for the CP, confirming it as a distinct, novel species. 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.

Genomic characterization of ALRSV and MLRSV provided here confirms their identification as nepoviruses based on the biological characterization previously available [2-6] and, especially ALRSV as a member of subgroup C [6]. In particular, phylogenetic analyses together with the long highly conserved 3' UTRs between the genomic RNAs strengthen ALRSV and MLRSV in subgroup C, in contrast to PMVA in subgroup A with short 3' UTRs [1]. Unlike the report of a satellite RNA associated with MLRSV [3, 4], we did not identify any satellite RNA either with MLRSV or ALRSV. Due to the fact that only one region identity (CP) in ALRSV and MLRSV meets the species demarcation criteria [1], the assignment of ALRSV and MLRSV to two distinct species is not quite straightforward, particularly if we consider their biological properties. Both viruses were found in the south of France, displaying some similar symptoms when grafted on the

same host, as symptoms of leaf enation and russeting on cherry (cultivar Bing), chlorotic spots on leaves, shortened internodes and stunting on GF305 [6]. In addition, when ALRSV was grafted on *P. cerasifera* rootstock, the natural host of MLRSV, similar symptoms induced by MLRSV appeared, as stope growth, and shortened internodes [6]. When it comes to their relative herbaceous host range [2-6], both ALRSV and MLRSV displayed the systemic typical ring pattern of nepoviruses on *Nicotiana tabacum,* as well as necrotic local and systemic lesions when inoculated to *Chenopodium quinoa* and *C. murale.* Whereas both asymptomatic in *Vigna sinensis,* they show differential behavior on *Cucumis sativus,* in which ALRSV induces no symptoms, while MLRSV causes a systemic mosaic, and on *Nicotiana clevelandii* (systemic necrotic mosaic for ALRSV and asymptomatic infection for MLRSV). Previous biological investigations have shown the lack of serological relationship of ALRSV with any of fruit tree nepoviruses, including MLRSV [6], which is not surprising regarding the 43% aa distance in their CP, as determined in this study. Moreover, the relatively low identity between their RNA1 (80% nt identity, 89 % aa) supports the absence of reassortment between ALRSV and MLRSV, which is one of the species discrimination criteria in the family *Secoviridae* [1]. Altogether, considering that not all the species discrimination criteria should simultaneously be met [1], we propose that ALRSV and MLRSV more likely should be considered as two distinct species. Finally, the genome availability of these three *Prunus*-infecting nepoviruses, including the novel PMVA, will assist to develop diagnostic assays useful for downstream epidemiology studies.

Acknowledgements

The authors thank the GeT-PlaGe Platform (GenoToul, INRAE, Toulouse, France) for high-throughput sequencing.

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. 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. https://doi.org/10.1099/jgv.0.000779
2. Dunez J, Delbos R, Desvignes JC, Marenaud C, Kuszala J and Vuittenez A (1971) Evidence of a ringspot type virus on Prunus cerasifera (myrobolan B). Annales de Phytopathologie HS:117–128.
3. Fritsch C, Koening I, Murant AF, Raschké 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) Myrobolan latent ringspot virus, a bipartite genome virus and a strain of tomato black ring virus. Acta Hortic 67:324-324. https://doi.org/10.17660/ActaHortic.1976.67.42.
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. https://doi.org/10.1023/A:1011274610537.

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 (2022) The expanding menagerie of Prunus-infecting luteoviruses. Phytopathology. https://doi.org/10.1094/PHYTO-06-22-0203-R.
8. Martin, D. P., Murrell, B., Golden, M., Khoosal, A., Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. Virus evolution, 1(1).
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 Online First. https://doi.org/10.1007/s00705-022-05564-x.

Figure legends

Fig 1. Genome organization of the three *Prunus*-infecting nepoviruses apricot latent ringspot virus (ALRSV) (**a**), mirobolan latent ringspot virus (MLRSV) (**b**), prunus mira virus A (PMVA) (**c**). The predicted cleavage sites and their positions are shown by black triangles with annotated dipeptides (? indicates the candidate dipeptides). Predicted peptides are separated by dash lines. The nucleotide positions of 3' and 5' ends are marked on the straight line representing untranslated regions. The conserved motifs of helicase, RNA-dependent RNA polymerase (RdRp), and coat protein (CP) and their relative positions are shown. Co-Pro: protease cofactor; NTB: nucleotide triphosphate binding helicase; VPg: viral genome-linked protein; Pro: protease; Pol: polymerase; MP: movement protein; A(n) polyA tail.

Fig. 2. Neighbour-joining phylogenetic tree generated based on the coat protein sequence 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 references 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. Number of trimmed reads, average coverage and mapped reads percent to both genome segments of ALRSV, MLRSV and PMVA

Virus	Segment	Number of mapped reads to contigs	Percentage of the mapped read	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

Table 2. Pairwise comparison between ALRSV and MLRSV genomic regions

Segment	Full genome nt identity	Polyprotein aa identity	5’ UTR nt identity	3’ UTR identity
RNA1	80%	84%	56%	86%
RNA2	74%	70%	56%	86%

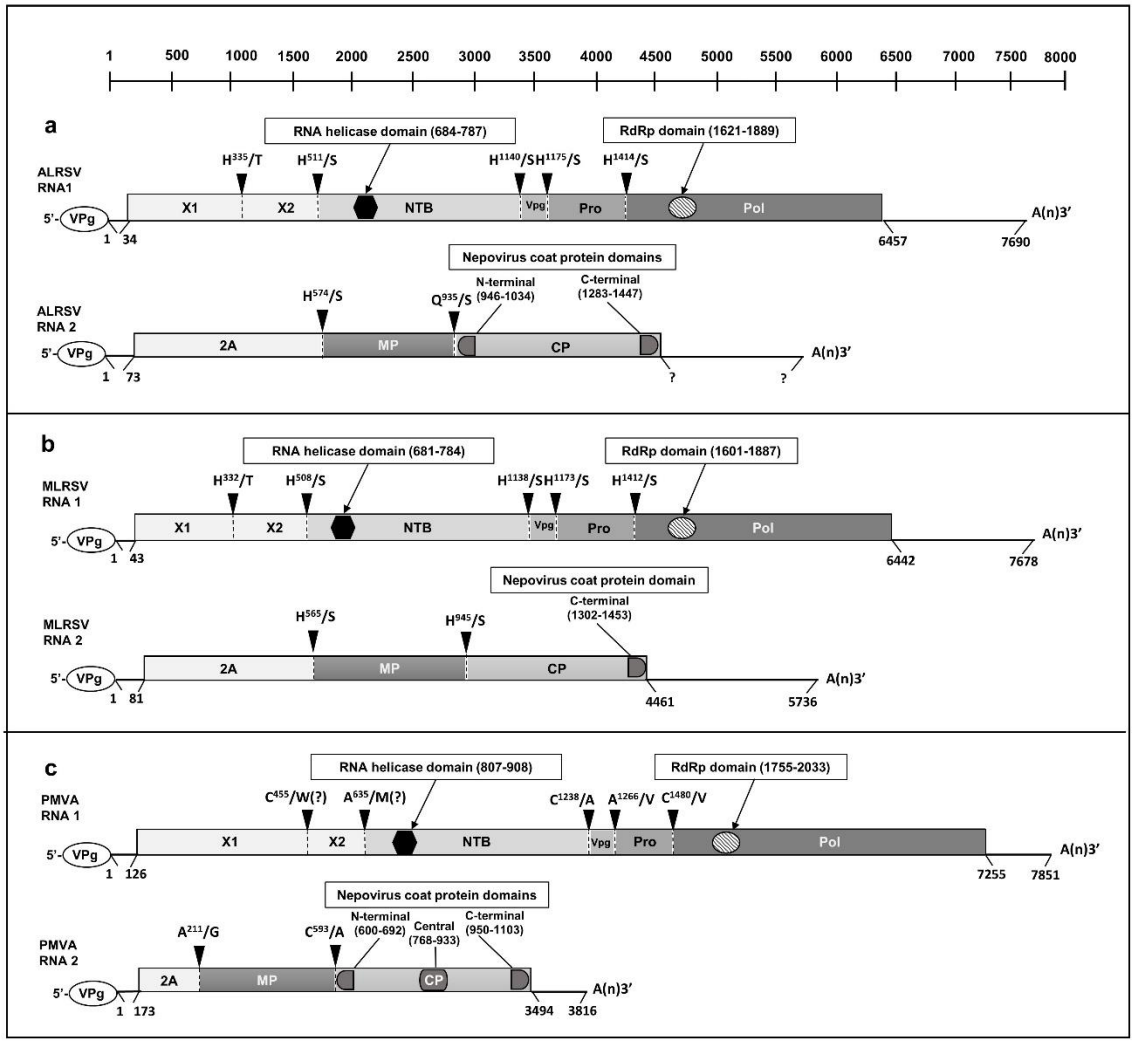

1 500 1000 1500 2000 2500 3000 3500 4000 4500 5000 5500 6000 6500 7000 7500 8000
a
RNA helicase domain (684-787)
RdRp domain (1621-1889)
ALRSV RNA1
H335/T H511/S H1140/S H1175/S H1414/S
5'- VPg X1 X2 NTB Vpg Pro Pol A(n)3'
1 34 6457 7690
Nepovirus coat protein domains
N-terminal (946-1034)
C-terminal (1283-1447)
ALRSV RNA 2
H574/S Q935/S
5'- VPg 2A MP CP A(n)3'
1 73 ? ?
b
RNA helicase domain (681-784)
RdRp domain (1601-1887)
MLRSV RNA 1
H332/T H508/S H1138/S H1173/S H1412/S
5'- VPg X1 X2 NTB Vpg Pro Pol A(n)3'
1 43 6442 7678
Nepovirus coat protein domain
C-terminal (1302-1453)
MLRSV RNA 2
H565/S H945/S
5'- VPg 2A MP CP A(n)3'
1 81 4461 5736
c
RNA helicase domain (807-908)
RdRp domain (1755-2033)
PMVA RNA 1
C455/W(?) A635/M(?) C1238/A A1266/V C1480/V
5'- VPg X1 X2 NTB Vpg Pro Pol A(n)3'
1 126 7255 7851
Nepovirus coat protein domains
N-terminal (600-692)
Central (768-933)
C-terminal (950-1103)
PMVA RNA 2
A211/G C593/A
5'- VPg 2A MP CP A(n)3'
1 173 3494 3816

Fig. 1

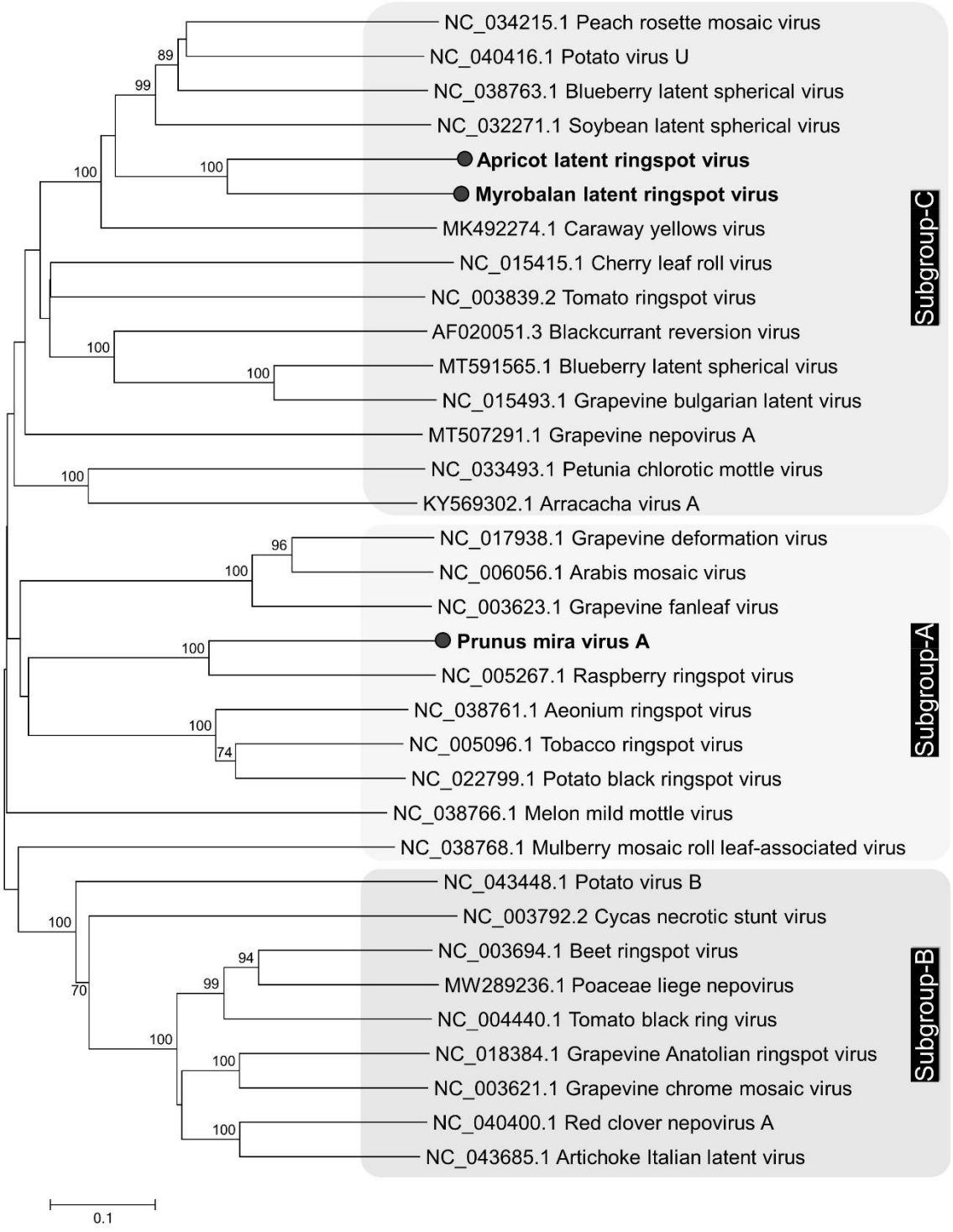

NC_034215.1 Peach rosette mosaic virus
NC_040416.1 Potato virus U
NC_038763.1 Blueberry latent spherical virus
NC_032271.1 Soybean latent spherical virus
Apricot latent ringspot virus
Myrobalan latent ringspot virus
MK492274.1 Caraway yellows virus
NC_015415.1 Cherry leaf roll virus
NC_003839.2 Tomato ringspot virus
AF020051.3 Blackcurrant reversion virus
MT591565.1 Blueberry latent spherical virus
NC_015493.1 Grapevine bulgarian latent virus
MT507291.1 Grapevine nepovirus A
NC_033493.1 Petunia chlorotic mottle virus
KY569302.1 Arracacha virus A
Subgroup-C
NC_017938.1 Grapevine deformation virus
NC_006056.1 Arabis mosaic virus
NC_003623.1 Grapevine fanleaf virus
Prunus mira virus A
NC_005267.1 Raspberry ringspot virus
NC_038761.1 Aeonium ringspot virus
NC_005096.1 Tobacco ringspot virus
NC_022799.1 Potato black ringspot virus
NC_038766.1 Melon mild mottle virus
NC_038768.1 Mulberry mosaic roll leaf-associated virus
Subgroup-A
NC_043448.1 Potato virus B
NC_003792.2 Cycas necrotic stunt virus
NC_003694.1 Beet ringspot virus
MW289236.1 Poaceae liege nepovirus
NC_004440.1 Tomato black ring virus
NC_018384.1 Grapevine Anatolian ringspot virus
NC_003621.1 Grapevine chrome mosaic virus
NC_040400.1 Red clover nepovirus A
NC_043685.1 Artichoke Italian latent virus
Subgroup-B
0.1

Fig. 2

GENERAL DISCUSSION AND CONCLUSION

General discussion

The primary purpose of my thesis was simply to explore the *Prunus* virome and identify viruses that have not yet been discovered. To get insight into the viral community in *Prunus*, I chose to use the HTS-based viral enrichment method, by purifying and sequencing the dsRNA from the plants, used as a proxy of infection by RNA viruses. We have developed and applied an analysis strategy for metagenomic data, based on dsRNA random amplification in stone fruit virome characterization. To the best of our knowledge, this is the largest survey of *Prunus* virome in terms of number of samples and diverse range of *Prunus* species and varieties.

The main outcomes of this dissertation have been thoroughly discussed in each chapter. Here are the latest reflections on what we know about virome of *Prunus*, how comparable is our study with relative ones, the impact of HTS in plant virus diagnostics, and the impact of virome characterization of this thesis.

1- To what extent is our knowledge of *Prunus* virome complete, and how much is still to be explored?

The cost of sequencing and the speed (throughput capacity) of sequencing have long been prohibitive for metagenomic analyses. With current advancements in HTS, it's not really an obstacle anymore and is becoming increasingly affordable. As a result, HTS has become more widely applied in research, including virus discovery projects. Overall, in the present study, 22 known viruses, two viroids and seven novel *Prunus* viruses (two from datamining analyses) were identified as composing the virome of *Prunus* species. However, we could wonder at which extend our virome description is exhaustive. In other words, did we miss viruses and viroids in our investigation, how does it compared with equivalent studies on *Prunus* virome discovery and what would be the best samples to choose in order to get more insights into the virome description of *Prunus* species (including the discovery of novel viruses)?

1.1 Sampling strategy

Prior to HTS becoming popular in plant virology, our knowledge was limited to economically important cultivated hosts and viruses that cause symptoms in their host mainly due to collecting the symptomatic leaves (Wren *et al.*, 2006). Therefore, latent viruses or inducing asymptomatic infections have been neglected for a long time in the viral community and their contribution to the holobiont has been under-estimated (Bordenstein & Theis, 2015; Roossinck 2019). This might lead to the spread of latent viruses that could pose symptoms in another plant species or cultivar. In this study, random samples were taken regardless of the symptoms. By randomly collecting the samples, we were supposed to be able to unbiasedly look for viruses in *Prunus*.

1.2 HTS strategy

Due to the number of samples to deal with and considering the low concentration of fruit tree viruses in *Prunus* trees, dsRNA enrichment approach was chosen to carry out the survey. Viral enrichment of dsRNA (randomly amplified) has been applied as the main methodological approach in almost all HTS analyses performed in this thesis, except for the completion of the genome of the two nepoviruses ALRSV and MLRSV (Chapter IV), which were analyzed by rRNA-depleted RNAseq, in order to improve the coverage and genome assembly step, as explained in the introduction.

However, it has been shown that HTS targeting dsRNA often failed to detect DNA viruses, albeit one DNA virus has been identified to date (Al Rwahnih *et al.*, 2018). In addition, single stranded negative-sense viruses are also difficult to detect by this dsRNA approach (Bejerman *et al.*, 2020). We did not identify viroids, but in peach trees (peach latent moaic viroid, PLMVd) and only in one apricot accession (hop stunt viroid, HSVd).

The low concentration of fruit tree viruses and their uneven distribution can be a limit for any detection approach including HTS and reduce the sensitivity of the assays. By pooling multiple samples in an HTS library, the depth of sequencing and coverage will be decreased leading to incorrect conclusions of the false negative signals as the absence of the virus (Maclot *et al*., 2020; Visser *et al.*, 2016). Interestingly, 14 apricot accessions shared with the samples we analyzed were submitted to mRNAseq experiment (Bui, personal communication),

allowing to compare the virome determined by dsRNAseq and mRNAseq. In 5/14 samples, the virome composition was identical between the two approaches. In other five samples, dsRNA strategy allows the detection of additional viruses in comparison to mRNAseq. The viruses detected only by dsRNAseq were mainly identified as viruses lacking a polyA tail, which is in agreement with the limit of the mRNAseq strategy to identify plant viruses. Finally, in the remaining cases, RNAseq identified the viroid HSVd (3/4 cases), which is known to be detected with difficulty by dsRNAseq. Interestingly, no DNA virus was detected by mRNAseq. Nevertheless, this comparison and what we can infer from it is rather limited. Indeed, a lot of other factors such as the collect period of the samples, the uneven distribution of viruses etc… could also explain some discrepancies between the two approaches.

1.3 How comparable is our study with relative ones?

This study is one of the largest efforts have been made to explore *Prunus* virome to date. Beside the wild samples collected from natural orchards and private gardens, the main reason to analyze *Prunus* samples from a germplasm collection was the fact that BRCs gather and preserve the existing diversity of plant varieties originating from various countries; in the *Prunus* INRAE BRC, 250 accessions representing 247 different varieties originating from more than 30 countries were investigated. This would consider as a quite comprehensive study on the virome of *Prunus* species. Surprisingly the number of the new viruses reported in the current study with regard to the number of samples analyzed is very small. In addition to the three luteoviruses MaLV, AlLV1, and ChALV which were identified in single samples collected from a *P. mume* tree collected in Japan, an almond tree collected in Turkey, and a cherry tree from Czech Republic respectively, we identified only two new viruses from BRC samples, the luteovirus PaLV2 in *P. persica*, and the cheravirus AWPV in *P. armeniaca*. No novel virus has been found in plum, sweet cherry and sour cherry. Moreover, only two new viruses, the luteovirus PmaLV in *P. mahaleb* and an isolate of AWPV isolate in *P. brigantina* both collected in France, were identified from 233 wild *Prunus* samples originated from France, Kazakhstan, Kirghizstan and Azerbaijan. I should also point out that we did not take into account novel viruses recovered from datamining, *i.e*, PhaLV, the new luteovirus in *P. humilis* dataset, and the PMAV, the novel nepovirus identified in *P. mira* dataset.

Two similar studies on peach virome were recently conducted in the USA (Dias *et al.,* 2022) and Hungary (Barath *et al.,* 2022). The viromes of Hungarian peach trees from germplasm were determined using vsiRNA method. Ten different *Prunus* cultivars were collected from isolator house, in addition to two cultivars from both isolator house and open field (aphid proof) isolator from Hungarian germplasms. Similar to our finding, NSPaV and PaLV were frequently (in *ca* 50% of tested samples) identified in Hungarian peaches in both isolator house and open field. Likely PLMVd infection rate was observed very high. Surprisingly from results of our European survey (Chapter II), PaLV2, the new luteovirus identified in peach in our study was not present in screened peach samples. In the US study, a total of 100 peach trees from 14 orchards of Tennessee (different cultivars) were analyzed by RNA seq. Nine known viruses have been reported as the viruses that constitute the peach virome including CGRMV, CNRMV, PNRSV, grapevine associated Ilarvirus (GaIV), turnip vein-clearing virus (TVCV), white clover mosaic virus (WClMV), tomato mosaic virus (ToMV), nectarine virus M (NeVM), tomato ringspot virus (ToRSV). However, the presence of GaIV, TVCV and CGRMV was not confirmed by RT-PCR and there is no data on how the presence of the ToMV has been validated. Consequently, some unsubstantiated and unvalidated data render these peach virome reports unreliable. No novel peach viruses have been either reported from this study. Although in both studies there are some first-reports of known viruses [PaLV: first report in Hungary, NeVM: first report in peach which is the same plant species as nectarine (*P. persica*)], they did not find any new virus and therefore the virome of peach was not expanded.

Altogether, we might be able to conclude that we are probably reaching the plateau of the *Prunus* virus discovery. There may still be viruses to be discovered, but it is more likely that the number of these unrevealed viruses is not so high.

1.4 What is the best strategy to further explore the *Prunus* virome?

1.4.1 Scanning biological resource centers

Because the newly established commercial orchards are using certified plant materials, there is very little chance of finding a new virus. One of the interesting places to be considered

is germplasms, because much efforts have already been made to search for and gather different varieties and various indigenous, wild and cultivated plant species. In particular, the most interesting accessions in a given germplasm would be ancient local varieties native to the country of interest which are not found in the other germplasms across the world.

1.4.2 Scanning the diversification center of Prunus

Apart from germplasms, traditional small orchards with locally grown varieties that are managed in a low-intensity basis might be the most efficient virus discovery strategy, particularly in the center of the *Prunus* diversification, located in Middle and South Asia along with the countries adjacent to the Caspian Sea (Blando and Oomah 2019; Janick 2005). However, our results of comparing the wild and cultivated viromes illustrated that the virome of cultivated crops is notably enriched, with 27 viruses and two viroids in cultivated *Prunus* in comparison with only 11 viruses in wild ones. New viruses may still be found in wild species that have never been analyzed, but the number of viruses most likely won't be significant.

1.4.3 Data mining: a new era in virus discovery

The approaches used to identify viruses have been fundamentally changed from the first discovery of viruses. Nowadays, HTS technologies being used are now producing a massive amount of sequencing data. As HTS-derived sequences are incrementally growing, plant virology researchers take advantage of this precious resource to dig the publicly available sequence read archives (SRA) databases. Therefore, virus discovery has started to undergo another transformation from processing biological data toward re-using freely available sequencing data. As a result, the main advantage of data mining over the conventional virus discovery methods that cannot be overestimated is the discovery of both known and novel viral sequences as a "by-product" of sequencing host without any wet-lab effort at a global scale (Edgar *et al.*, 2022). As an example, Sidharthan *et al.* (2021) recently concentrated on searching for new viruses in endangered plant species, leading to the discovery of nine putative novel viruses in transcriptomic datasets.

However, the traditional data mining method requires enormous computational processing that can be time-consuming and expensive. Recently, Edgar *et al.*, (2022)

developed a free cloud computing infrastructure for sequence alignment called Serratus (https://github.com/ababaian/serratus), which revolutionized the discovery of RNA viruses (the method used in Chapter II for discovery of PhaLV). In fact, analysis of this global scale data that would not have been possible with conventional data mining methods is now feasible. It has to be considered that Serratus, which is based on the homology of an RdRp conserved motif, is unable to predict DNA viruses, viroids and viruses with no significant homology with RdRp motif; hence, the conventional method of data mining (download the SRA data with no preference, which was the strategy for identifying PMVA in Chapter IV) may be implemented for that purpose. It is worth noting that, although the development of data mining methods helps to uncover new viruses, they are still homology-based approaches, therefore, incapable of processing the vast sequence data that remain unclassified (the so-called dark matter). As mentioned above, we may reach the plateau in *Prunus* virus discovery using the homology-based approaches, which illustrates the importance of a great effort to develop strategies that are able to distinguish the viral sequences from the other entities regardless of the identity level of the sequence with existing viruses in the databases.

Nevertheless, we should not forget that, as described in Massart *et al.* (2020), the presence of a virus has to be biologically confirmed, and this might be challenging due to the lack of plant materials. Furthermore, a virus derived from *Prunus* transcriptome data is not certainly a *Prunus*-infecting virus; it may be due to contaminations either in the lab preparing the materials or in the sequencing platform (Laurence *et al.*, 2014; Schmieder & Edwards, 2011). It may be also a virus infecting a *Prunus*-associated organism, such as fungal pathogen (Hily *et al.*, 2018). This is another data mining challenge unless we can contact the laboratory where the sequences were generated and request the plant material, if it remains, to validate the presence of the virus.

Beyond the discovery of novel viruses, one of the applications of data mining is related to its potential for exploring the genetic diversity of known viruses worldwide by revealing multiple virus isolates in different hosts or from various geographical locations, which brings additional information to the lab-derived sequences. For instance, Hily *et al.* (2020) performed a systemic datamining on SRA data of grapevine. In addition to their HTS data, 50 complete or nearly complete genomes were thus obtained for grapevine Pinot gris virus and grapevine

berry inner necrosis virus, providing new insights into their evolutionary history and epidemiological success. Moreover, by identification of divergent variants of a known virus, data mining can assist to develop a more accurate diagnosis by providing additional genetic diversity information (*e.g.* Nourinejhad-Zarghani *et al.,* 2018).

Overall, data mining offers a very economic and eco-friendly promising strategy to fill the gaps in our wet-lab-derived knowledge. In addition, new data mining tools like Serratus have opened up a new era in exploring existing sequencing data in a few seconds.

2- Impact of HTS in plant virus diagnostics

Several recent reviews have been devoted to the application of HTS approaches in plant health domain, in particular as a diagnostic tool (*e.g.* Adam *et al.,* 2018; Kutnjak *et al.,* 2021; Mehetre *et al.,* 2021; Villamor *et al.,* 2019). As I don't want to paraphrase these studies, the sole aspects that have been less addressed will be discussed here. As repeatedly, time and price are the major obstacles to HTS application in daily diagnosis. Although Illumina sequencing cost is progressively decreasing and pooling strategies can reduce the price, it is not affordable for many laboratories to include HTS in the routine diagnostic assay. In addition, high-quality materials are required for sequencing, which is highly dependent on laboratory set-up. Therefore, when a well-characterized known virus is subject to large-scale surveillance, it is more reasonable to use PCR and ELISA. However, the ability of HTS approach to address one of the only properties absolutely conserved between all viruses (*i.e.* the encoding of genetic information in the form of nucleic acids), and therefore potentially representing an absolutely generalist technique, capable of detecting any viral agent without a priori, makes it a suitable approach to be implemented for the validation of the health status of valuable plant material, such as propagation clone heads, genetic resources... In addition, some HTS approaches such as RNAseq or siRNAseq are able to simultaneously identify not only viruses, but also other pathogens, including fungi, bacteria, and phytoplasmas, in one trial, even there are still technical and bio-informatic challenges to be meet.

Recent advances in the fourth-generation sequencing, Oxford Nanopore technology (MinION), hold great promise in plant virus diagnosis (Bronzato Badial *et al.,* 2018; Liefting

et al., 2021). It offers a lower price, rapid and long read sequencing (complete viral genome) with real-time data delivery that could be installed in every laboratory. The drawback of using MinION is the high error rate (Filloux *et al.,* 2018) of sequencing reads which might not crucial in plant virus diagnostic due to enough genome coverage. In addition, the high error rate may be resolved by choosing the appropriate MinION device and enough sequencing depth (Pecman *et al.,* 2022). Implementing MinION in the quarantine program may reduce the worklab and speed up the entire phytosanitary certification process.

However, there are challenges to applying HTS, even though a guideline about applying HTS as a plant pest diagnostic test in laboratories was published by the European and Mediterranean Plant Protection Organization (EPPO) through the VALITEST project: Shaping the future of plant pest diagnostics in the EU) (Lebas *et al.,* 2022; Massart *et al*., 2022). As described in the introduction session, the sensitivity and specificity of each HTS approach differ as the main challenges of the HTS application for diagnostic purposes. To improve the sensitivity of the HTS for low titer viruses, the enrichment methods can be an alternative to non-enrichment ones, depending on the intended use of the HTS. The combination of classic techniques and HTS can also enhance the specificity of HTS. During this study, the presence of viruses was confirmed by RT-PCR to reduce false positives of certain suspicious viral sequences. Altogether, HTS-based approaches are becoming routine assays, particularly in quarantine programs, due to the advances in sequencing platforms and optimizing the process and material for sequencing (Soltani *et al*., 2021). However, we are a long way away from replacing conventional diagnostic methods with HTS.

3- Conclusion: Impacts of the *Prunus* virome characterization of this thesis

As already explained in the introduction, epidemiologic surveillance and risk assessment are essential components of a disease management program. Risk assessment defines the risks posed by viruses to crops or the environment and identifies control measures required to mitigate the viral disease risks using a range of strategies including quarantine programs, certification, or removal of infected plants. Moreover, epidemiological surveillance and modeling systems can predict the probability of a particular viral disease outbreak, but rely on the availability of risk factor data that are often partial or missing and therefore need to be

more extensively developed. In conclusion, all the disease management approaches rely on the early and accurate detection of viruses in plants as the basis for every successful program to respond to a more significant number of threats posed by emerging and re-emerging viruses. In this context, I discovered some novel viruses in *Prunus* species to provide essential data on the viral species infecting *Prunus* and define the stone fruit virome using cutting-edge technology of high throughput sequencing, identified viruses that had remained unknown, then validated and designed the detection assays targeting newly discovered viruses. In addition, establishing a link between a virus presence and the potential symptoms provides a data base for authority bodies and plant health organization to inform them whether they have to immediately take an action to avoid virus spread or not. Two new viruses identified in this study, PaLV2 and AWPV, have begun to be biologically characterized. However, more efforts need to fully characterize the rest of new viruses discovered in my thesis project, especially MaLV shown to infect *P. mume* but also *P. armeniaca* and *P. insica* and ChLVA found in cherry trees as well.

ANNEX I

Scanning of *Prunus* virome

Annex 1: Scanning of *Prunus* virome

1- Plant material and sampling

Fifty leaf samples of each *Prunus* species including *P. domestica*, *P. persica*, *P. armeniaca*, *P. cerasus*, and *P. avium* were collected regardless of the symptoms from different parts of the trees considering different height and canopy in INRAE *Prunus* BRC, (https://www6.bordeaux-aquitaine.inrae.fr/uea/Axes-de-Recherche/Centre-de-Ressources-Biologiques) which gathers traditional and modern varieties of French or foreign origin introduced between 1937 and 2010. Taken together, 250 *Prunus* accessions originated from more than 30 countries were thus sampled. Besides, 243 samples belonging to wild *Prunus* species such as, *P. brigantina*, *P. mahaleb*, *P. spinosa*, *P. serotina*, *P. cerasifera*, were collected from Azerbaijan, Kazakhstan, Kirgizstan and France between 2017 to 2021 (Table 1).

2- Methods

2-1. Double-stranded RNA extraction, HTS, bio-informatic and phylogenetic analyses

These methods have been already described in the Chapter II of this thesis.

2-2. RT-PCR confirmation:

Depending on the sequencing platform and every HTS run, we decided to set a threshold for mapped reads against the reference genome to discriminate the false positives from true ones. In cases with low mapped read numbers observed in the absence of any contigs of the relevant virus, an infection was therefore considered as suspected, and the presence of the virus was examined through PCR assays. The primers used to detect the viruses are listed in Table 2, and RT-PCR conditions are those of the references.

Table 1. List of the *Prunus* species investigated by HTS, number of samples, their origin and year of sampling.

Plant type	Species	Number of accessions	Origin	Collection year
Cultivated	*Prunus persica*	50	Germplasm[a], 11 countries	2019
	Prunus domestica	50	Germplasm[a], 18 countries	2019
	Prunus avium	50	Germplasm[a], 16 countries	2019
	Prunus cerasus	50	Germplasm[a], 14 countries	2019
	Prunus armeniaca	50	Germplasm[b], 17 countries	2019
Wild	Wild almond	16	Azerbaijan	2017
	Wild apricot	24	Kazakhstan, Kirgizstan	2017
	Prunus brigantina	15	Nevach, France	2017
	Prunus mahaleb	1	Aussois, France	2021
	Wild plum	5	Bordeaux, France	2020
	Prunus cerasifera	49	Bordeaux, France	2020
	Prunus spinosa	66	Bordeaux, France	2020
	Prunus spinosa	2	Aussois, France	2021
	Prunus serotina	24	Bordeaux, France	2020
	Prunus subcordata	3	Germplasm [a], France, USA	2019
	Prunus serrulata	8	Germplasm [a], 7 countries	2019
	Prunus maximowiczii	1	Germplasm [a], France	2019
	Prunus nipponica	3	Germplasm [a], Japan, Netherland	2019
	Prunus americana	1	Germplasm [a], Denmark	2019
	Prunus japonica	1	Germplasm [a], France	2019
	Prunus fontanesiana	2	Germplasm [a], France, Canada	2019
	Prunus conradinae	1	Germplasm [a], USA	2019
	Prunus sargentii	1	Germplasm [a], France	2019
	Prunus mahaleb	1	Germplasm [a], France	2019
	Prunus apetala	1	Germplasm [a], USA	2019
	Prunus cerasifera	6	Germplasm[a], 5 countries	2019
	Prunus hillieri	1	Germplasm [a], UK	2019
	Prunus juddii	1	Germplasm [a], UK	2019
	Prunus subhirtella	2	Germplasm [a], France, UK	2019
	Prunus dawyckensis	2	Germplasm [a], France	2019
	Prunus yedonsis	2	Germplasm [a], USA, Japan	2019
	Prunus schmittii	1	Germplasm [a], Netherland	2019
	Prunus pendora	1	Germplasm [a], UK	2019
	Prunus tomentosa	1	Germplasm [a], France	2019
	Prunus insica	1	Germplasm [a], Japan	2019

[a] *Prunus* INRAE Biological Resource Center (BRC Toulenne)

[b] *Prunus* INRAE Biological Resource Center (BRC Avignon)

Table 2. List of primers used for confirmation of HTS data by RT-PCR

Virus	Primer name	Primer sequence 5'-3'	Reference
ACLSV	A52	CAGACCCTTATTGAAGTCGAA	Candresse et al., 1995
	A53	GGCAACCCTGGAACAGA	
PNRSV	Ilar 1	TTCTAGCAGGTCTTCATCGA	Salem et al., 2004
	Ilar 2	CAACCGAGAGGTVTGGCA	
PDV	PDV1	TAGTGCAGGTTAACCAAAAGGAT	Spiegel et al., 1998
	PDV2	ATGGATGGGATGGATAAAATAGT	
LChV1	LChV1 upnest	GATTTAAARCGWTGGGTWGG	Katsiani et al., 2015
	LChv1 donest	CTTAAGTTTRWGTTGATCYTC	
CVA	CVA-F	GGAAATTTCTACATACATCA	Marais et al., 2012
	CVA-R	TAAAGATAATCCTCAAT	
ChVT	Tepo F1	GGCGGAGGTTCAGGGATGTG	Marais et al., 2020
	Tepo R1	TCATATCTAAGACATGTGAACAGC	
PLMVd	PLMvd-F	AACTGCAGTGCTCCGAATAGGGCAC	Jo et al., 2015
	PLMvd-R	CCCGATAGAAAGGCTAAGCACCTCG	

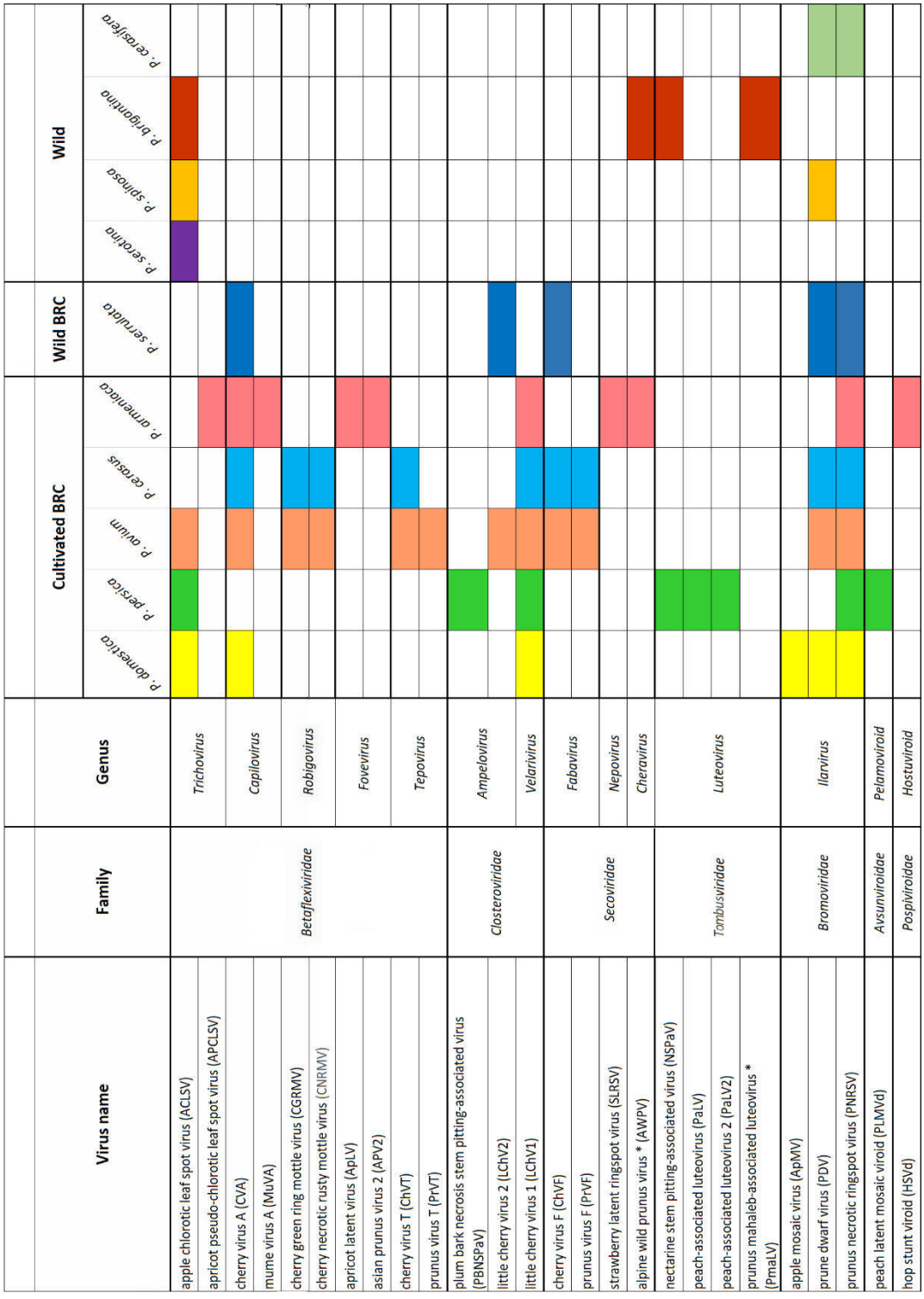
3- Results

The Table 3 summarizes the virome composition in the cultivated and wild species collected from the *Prunus* BRC as well as in the wild relatives. Globally, four viral families were detected in *Prunus* trees from BRC, including *Betaflexiviridae*, *Closteroviridae*, *Secoviridae*, *Bromoviridae*, and the two viroid families known so far, *Avsunviroidae* and *Popsiviroidae*, representing altogether 27 species, including three novel viruses, the cheravirus AWPV and the luteoviruses PmaLV and PaLV2, described in Chapter III and II. As shown in the Table 3, the virome of cultivated *Prunus* seem to be richer than the ones of wild *Prunus* species. The number of different viruses detected is comprised between 6 (plum) and 12 (sweet cherry), depending on the cultivated species considered. In contrast to the situation in cultivated species, the wild relative species host between 1 (*P. serotina*) and 4 (*P. brigantina*) viruses. No virus or viroid could be identified in wild almonds originated from Azerbaijan and wild apricots originated from Kazakhstan, Kirgizstan, despite the significant numbers of plants

analyzed, as well as other wild *Prunus* species that are listed in Table 4. From the virus species point of view, PNRSV was found the most prevalent virus, detected in all the five cultivated species and in two wild species (*P. cerasifera* and *P. serrulata*). Surprisingly, LChV1 was also found in the five cultivated species from BRC. ACLSV was detected in the same proportion in cultivated and wild species, as well as PDV (Table 3).

The following paragraphs are devoted to the analysis of the virome of each cultivated species.

Table 3. Virus and viroids identified in cultivated and wild *Prunus* (more than two samples /species)



* Novel viruses identified and characterized in this thesis

Table 4. List of *Prunus* species in which no virus or viroid could be identified**.**

Host	Number of samples
Wild almond	15
Wild apricot	24
Prunus cerasifera	15
Prunus nipponica	3
Prunus pseudocerasus	1
Prunus apetala	1
Prunus serotina	1
Prunus sargentii	1
Prunus hillieri	1
Prunus juddii	1
Prunus subhirtella	2
Prunus yedonsis	1
Prunus dawyckensis	1
Prunus tomentosa	1

3-1 Plum virome

The results showed that 24 out of 50 (48%) *P. domestica* accessions are infected by six known viruses including ACLSV, ApMV, CVA, LChV1, PDV, PNRSV. The prevalence of each virus and the combination of co-infecting viruses are shown in Fig. 1. CVA is the most prevalent virus, detected in 28% of the plums followed by LChV1 (14%) and PNRSV (12%).

Diversity analysis of the most prevalent viruses in plum

- **CVA**

CVA is a member of the genus *Capillovius* in the family *Betaflexiviridae* consisting of a single-stranded positive-sense RNA (+ssRNA) genome (Adams *et al.*, 2011). CVA was first

described from the sweet cherry in Germany (Jelkman, 1995). There have been some uncertainties on its epidemiology and biology, including its vector transmissibility and potential symptoms, which have not yet been studied. It can be transmitted by infected propagation materials through grafting (Gao *et al*., 2020). It has been reported with high frequency in cherry from different countries (Gao *et al*., 2020). Although CVA is less frequent in peach and plum in comparison with sweet and sour cherry (James and Jelkmann, 1998; Marais *et al.,* 2011; Sabanadzovic *et al.,* 2005; Svanella-Dumas *et al.,* 2005), it was found abundantly in plum in BRC collection; therefore, we decided to analyze its diversity. A shared region of HTS-derived contigs that had been assembled in 10 out of the 15 infected samples was selected [283 nt of the movement protein (MP) gene], aligned, and used to construct the phylogenetic tree shown in Fig. 2. In four cases, multiple variants in the same tree could be identified (shown in red in the Fig. 2). The CVA isolates are found in two clusters corresponding to the groups I and III, according to Gao et al. (2017), which might suggest multiple introduction events. The nucleotide diversity within isolates of group III is 1.7 % +/- 0.4, whereas it reaches and 12% +/- 1.8 between isolates from group I (only one variant) and group III. The very low genetic diversity of the majority of the samples raises questions about the likelihood of the vector transmissibility in the collection or the infection of rootstock. Even though, the rootstock infection cannot be excluded, the emergence of two very divergent variants is against that idea.

- **PNRSV**

It was first described in the USA from peach (Cochran and Hutchins 1941). PNRSV is a member of the genus *Ilarvirus* in the family *Bromoviridae*, which has three genomic +ssRNA (Bujarski *et al.,* 2012). It has been reported from all the stone fruits worldwide (Celik *et al.,* 2022). The symptoms occurrence and severity depend on different factors such as host species and cultivar, virus isolate, and sampling time. (Aparicio *et al.,* 1999; Gilmer *et al.,* 1976 ; Sokhandan-Bashir *et al.,* 2017). It can be transmitted through seed and pollen in combination with thrips feeding injuries, or pollinators (Pallas *et al.,* 2012). PNRSV isolates have been traditionally clustered into four groups, including PV96, PV32, CH30, and PE50 based on the CP or MP sequence (Aparicio *et al.,* 1999, Kinoti *et al.,* 2017).

Phylogenetic analysis (Fig. 3) based on the 298nt fragment of MP gene in the six isolates found in BRC shows that five isolates clustered together in the PV32 group, whereas the two variants found in the same tree (in red) are located in the PV96 group. Intra BRC isolates diversity was very low with 3.3%+/-0.8. Two variants were identified in a single plant that were placed in the distinct clades with 98.3% identity in 283 nt of MP, which may represent the genetic strains according to the definition of PNRSV strain by Kinoti *et al.* (2017), proposing a >97% sequence identity cut-off.

- **LChV1**

It was initially reported in Germany by Keim-Konrad and Jelkmann (1996) and distributed worldwide. It belongs to the genus *Velarivirus* in the *Closteroviridae* family. Although it has been associated with little cherry disease (Keim and Jelkmann 1996) in combination with LChV2, Kwanzan stunting (Matic *et al.,* 2009), and Shirofugen stuck disease (Candresse *et al.,* 2013) in cherry, it can be asymptomatic in other stone fruits (Katsiani *et al.,* 2015; Safarova *et al.,* 2017). Its genome consists of a +ssRNA. Apart from graft transmissibility, no vector has been identified to harbor the virus (Katsiani *et al.,* 2015). According to the latest study on the diversity of LChV1, the isolates are clustered in five distinct clades (Katsiani *et al.,* 2018). Analysis of the 429 nt of coat protein gene (CP) of LChV1 isolates found in BRC shows their clustering in group 3 (GP3) and group 5 (GP5). Intra Gp5 group diversity is 2.1%+/-0.5, and intraGp3 group diversity is 5.2%+/-1.6. Gp3/Gp5 intergroup diversity is 16.4%+/-1.6, corresponding to the overall genetic diversity of LChV1 across the world.

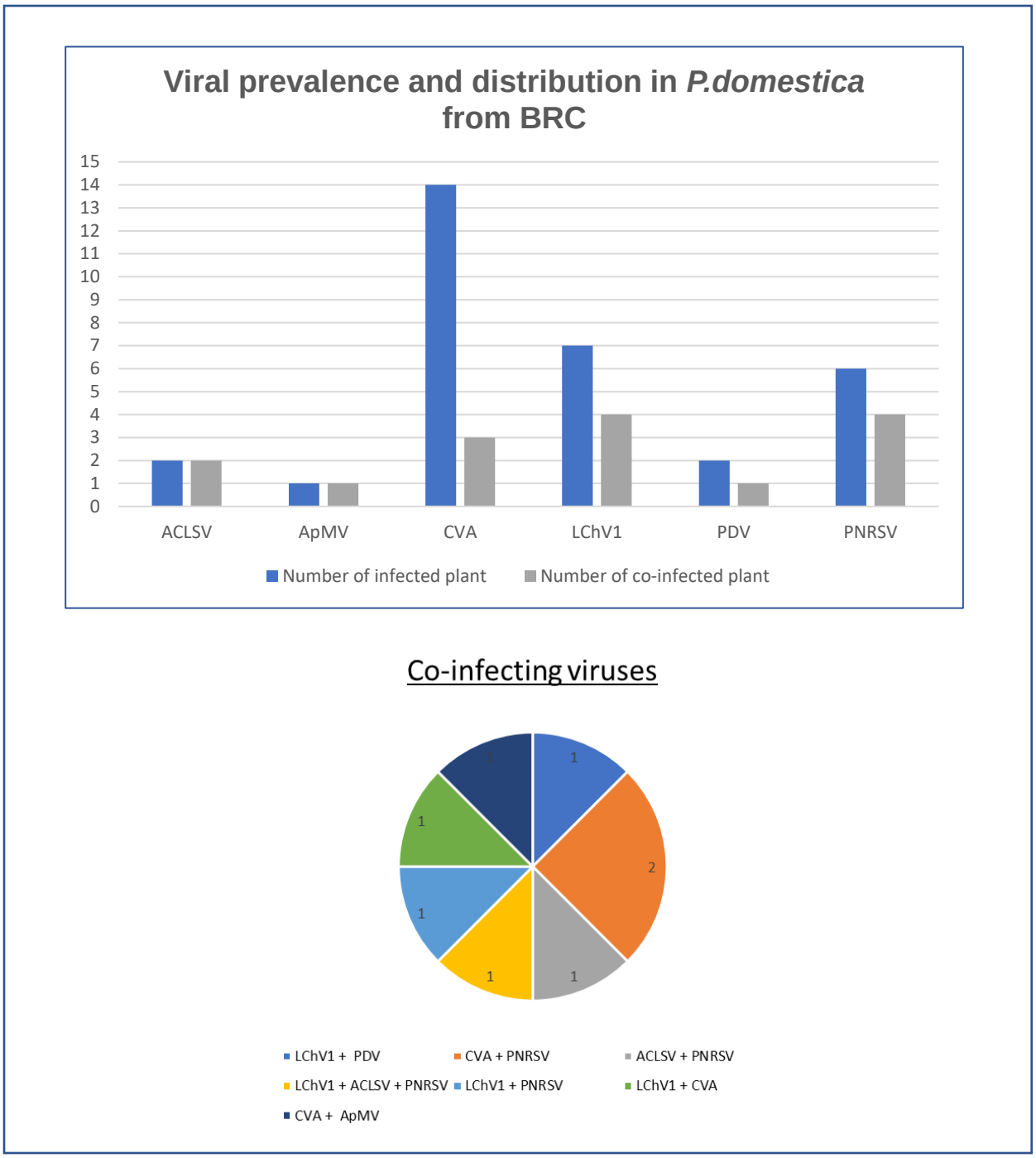


Fig 1. Prevalence of viruses found in plum in BRC collection and combination of the co-infecting viruses.

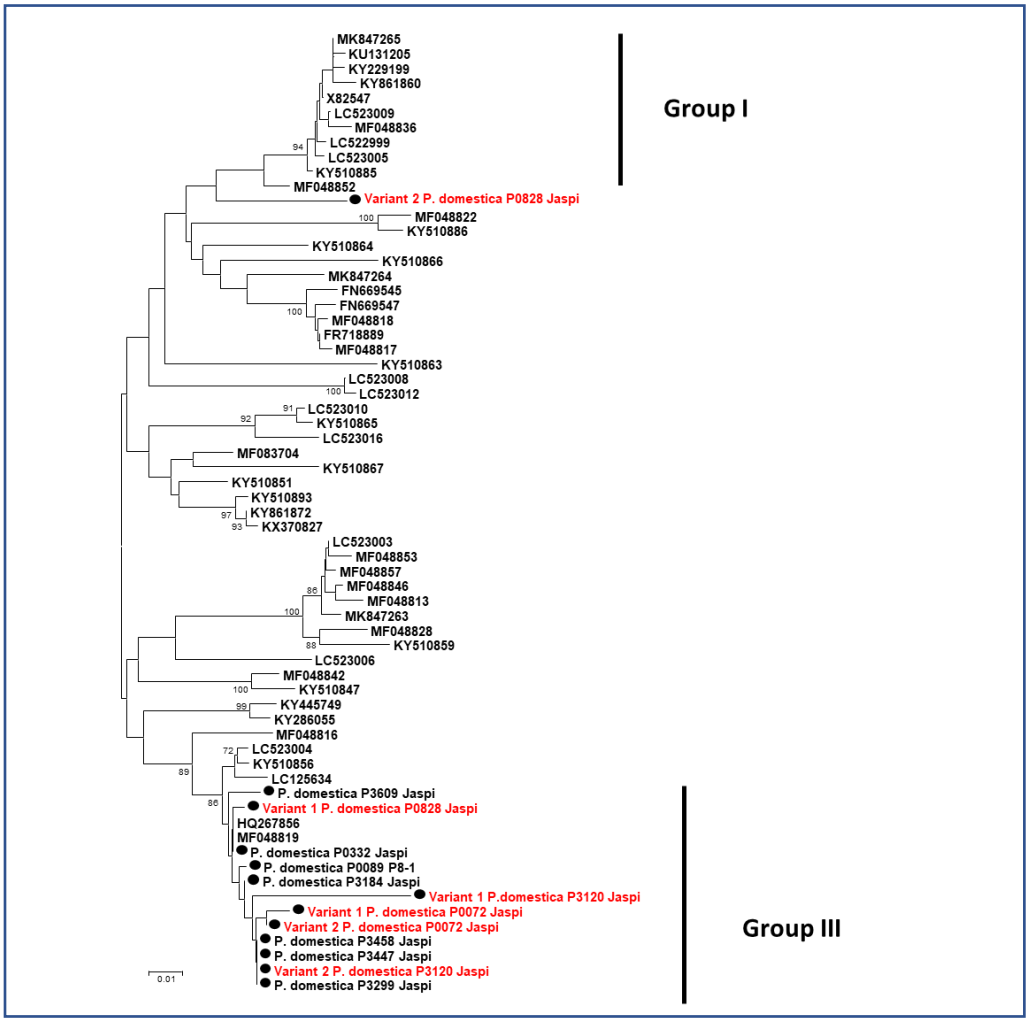


Fig 2. Phylogenetic tree based on the 283nt of the movement protein gene derived from HTS-based contigs of cherry virus A (CVA) in plum. Sequences generated in this study are shown by a black circle. Accession numbers of isolates from GenBank are indicated.

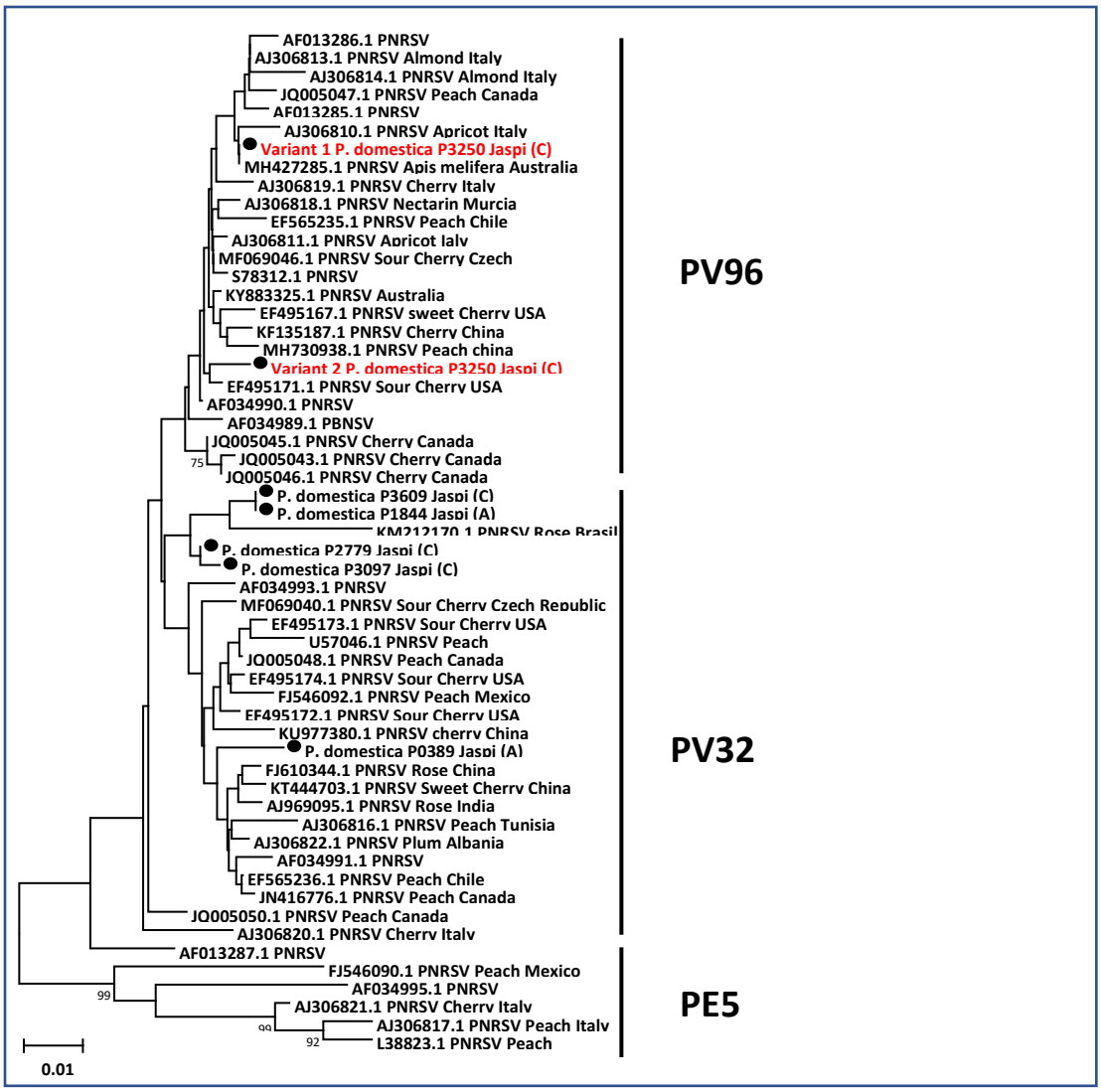


Fig 3. Phylogenetic tree based on the 298 nt of MP derived from HTS-based contigs of prunus necrotic ringspot virus (PNRSV) in plum. Sequences generated in this study are shown by a black circle. Accession numbers of isolates from GenBank are indicated

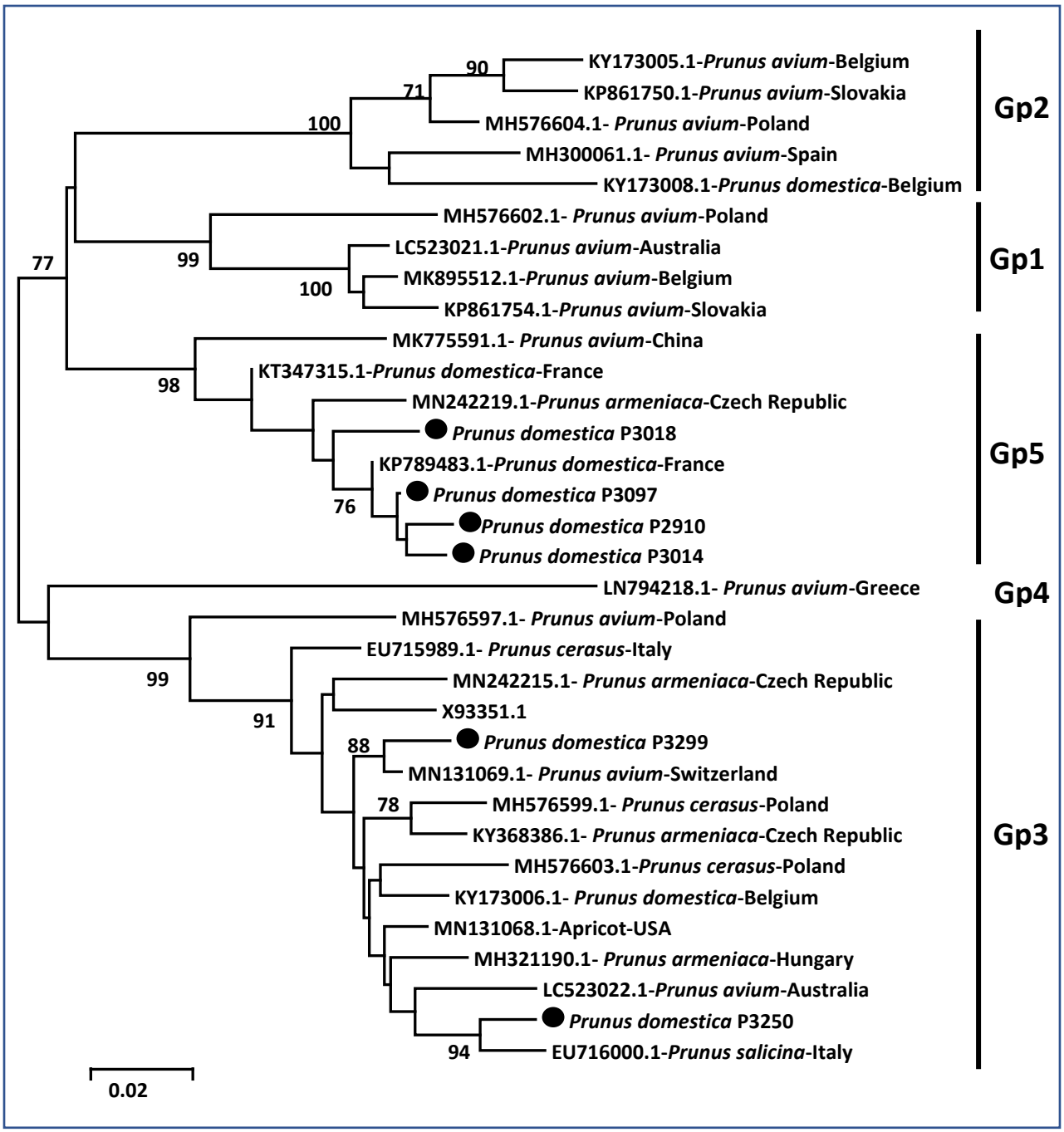


Fig 4. Phylogenetic tree based on the 429 nt of CP derived from HTS-based contigs of little cherry virus 1 in plum. Sequences generated in this study are shown by a black circle. Accession numbers of isolates from GenBank are indicated

3-3 Peach virome

All the 50 peach trees from the BRC collection analyzed were found to be infected by six known viruses (ACLSV, PBNSPaV, LChV1, NSPaV, PaLV, PNRSV), a known viroid (PLMVd), and a new luteovirus, PaLV2 (Fig. 5), which has been analyzed in detail in Chapter II. Only four samples were infected by a single virus, whereas 92% of the samples were co-infected with up to six viruses in 22 different combinations. The most frequent virus was NSPaV (96% of the tested peach trees), followed by PaLV2 (54%) and PaLV (38%), the tree luteoviruses that have been analyzed in detail in Chapter II.

Diversity analysis of the most prevalent viruses in peach

- **ACLSV**

ACLSV, a +ssRNA virus, is a member of the genus *Trichovirus* in the family *Betaflexiviridae* that was first isolated from US by Mink and Shay (1959). It has a worldwide distribution and can infect the members of the family *Rosaceae* (Katsiani *et al.,* 2014; Mathioudakis *et al.,* 2007; Yaegashi *et al.,* 2011). Symptoms may be severe in susceptible cultivars, from foliar malformation to declining trees, though ACLSV is generally latent. It is graft transmissible but no vector has been reported so far to transmit the virus (Katsiani *et al.,* 2014). Its high genetic diversity is driven by recombination is a main evolutionary process (Chen *et al*., 2014; Nakahara *et al*. 2011). The phylogram of 746 nt of the RdRp gene was constructed for 8/11 CRB isolates that are shown with black circles. Intra CRB isolates diversity was calculated to be 12.3 % +/- 0.8 Three variants of ACLSV were identified in a single tree that are indicated via red color clustering in three distinct clades with 13% overall nt diversity. This level of diversity is comparable with that found in other studies.

- **NSPaV, PaLV and PaLV2**

The genetic diversity of these three luteoviruses has already been examined in the chapter II. Regarding that they have been found abundantly in the collection with low genetic diversity may bring up the questions about their dissemination. Although their genetic diversity was low, the diversity of the European isolates was quite similar. Therefore, it is not easy to

determine the source of infection, whether it is the propagating material infected in germplasms or the origin of the plants.

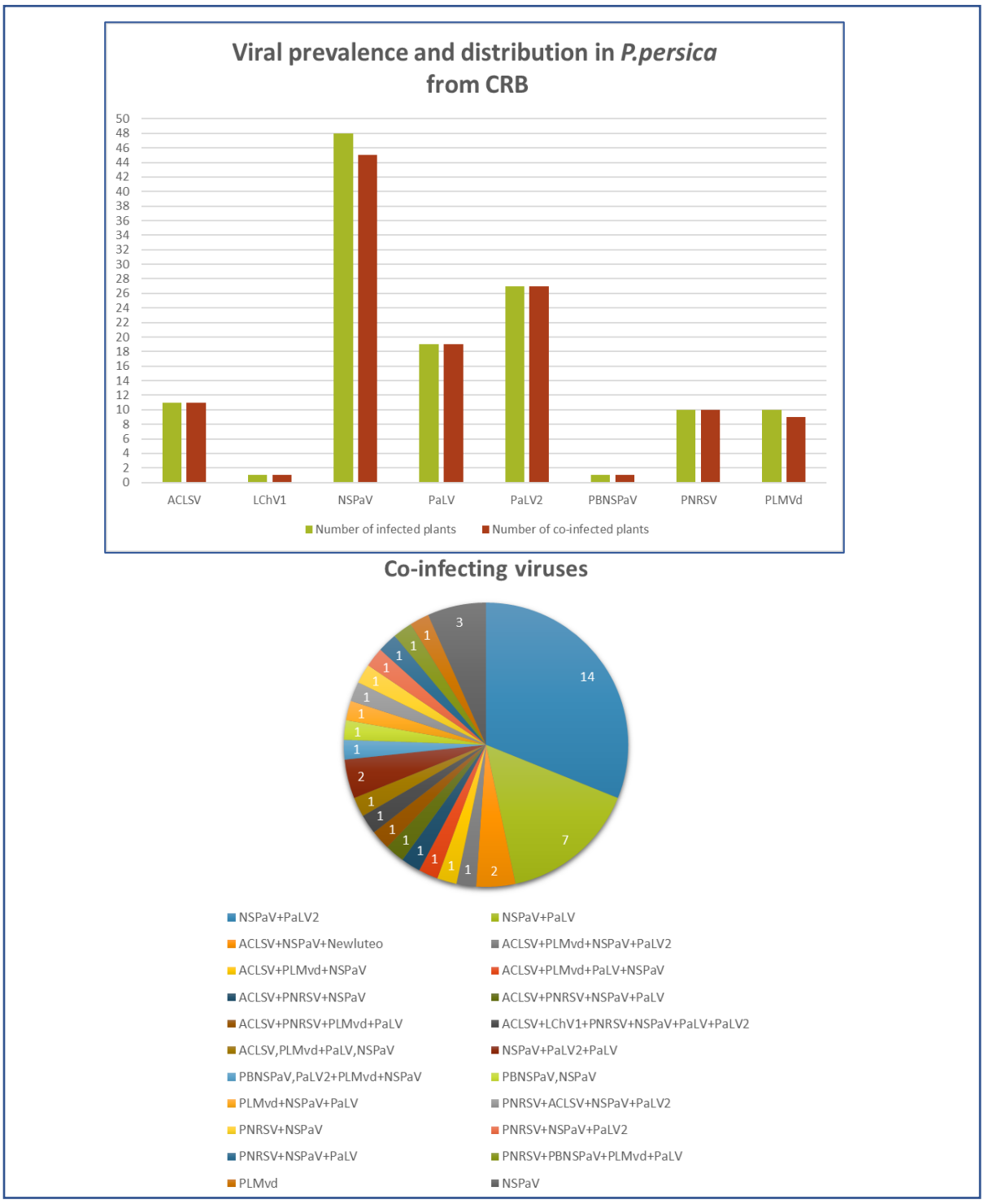


Fig 5. The prevalence of viruses found in peach in BRC collection and combination of the co-infecting viruses.

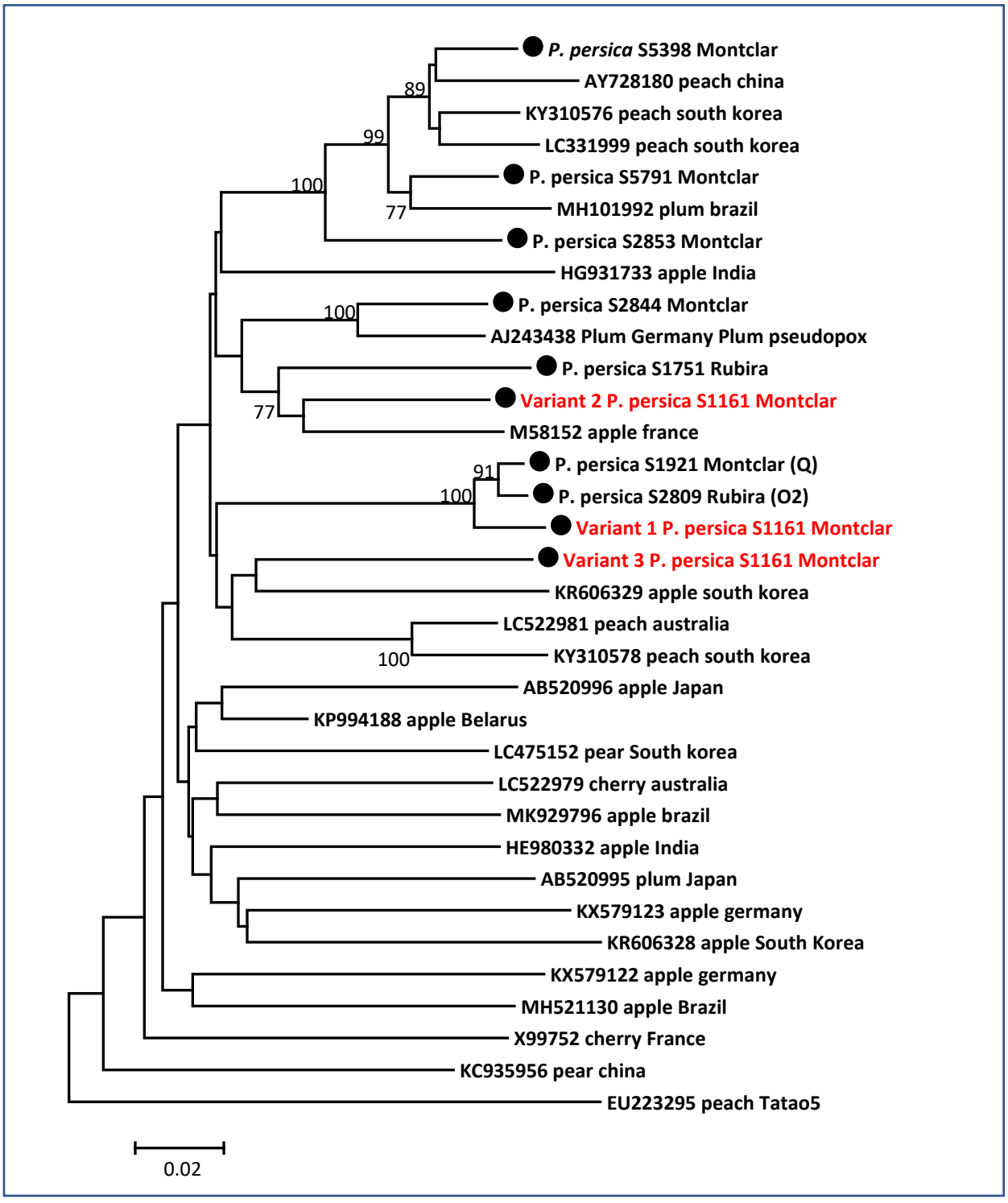


Fig. 6 Phylogenetic tree based on the 746 nt of RdRp derived from HTS-based contigs of apple chlorotic leaf spot virus in peach

3-4 Sweet cherry virome

As for the above *Prunus* species examined, most of the sweet cherry trees were infected (31/50, 62%). Twelve known viruses were detected: ACLSV, CVA, CGRMV, CNRMV, ChVT, PrVT, LChV1, LChV2, ChVF, PrVF, PDV, and PNRSV (Fig. 7). The most prevalent virus was CVA (30% of the tested samples) followed by PDV (28%) and LChV1 (18%). Sixteen combinations of co-infecting viruses have been observed in cherry trees.

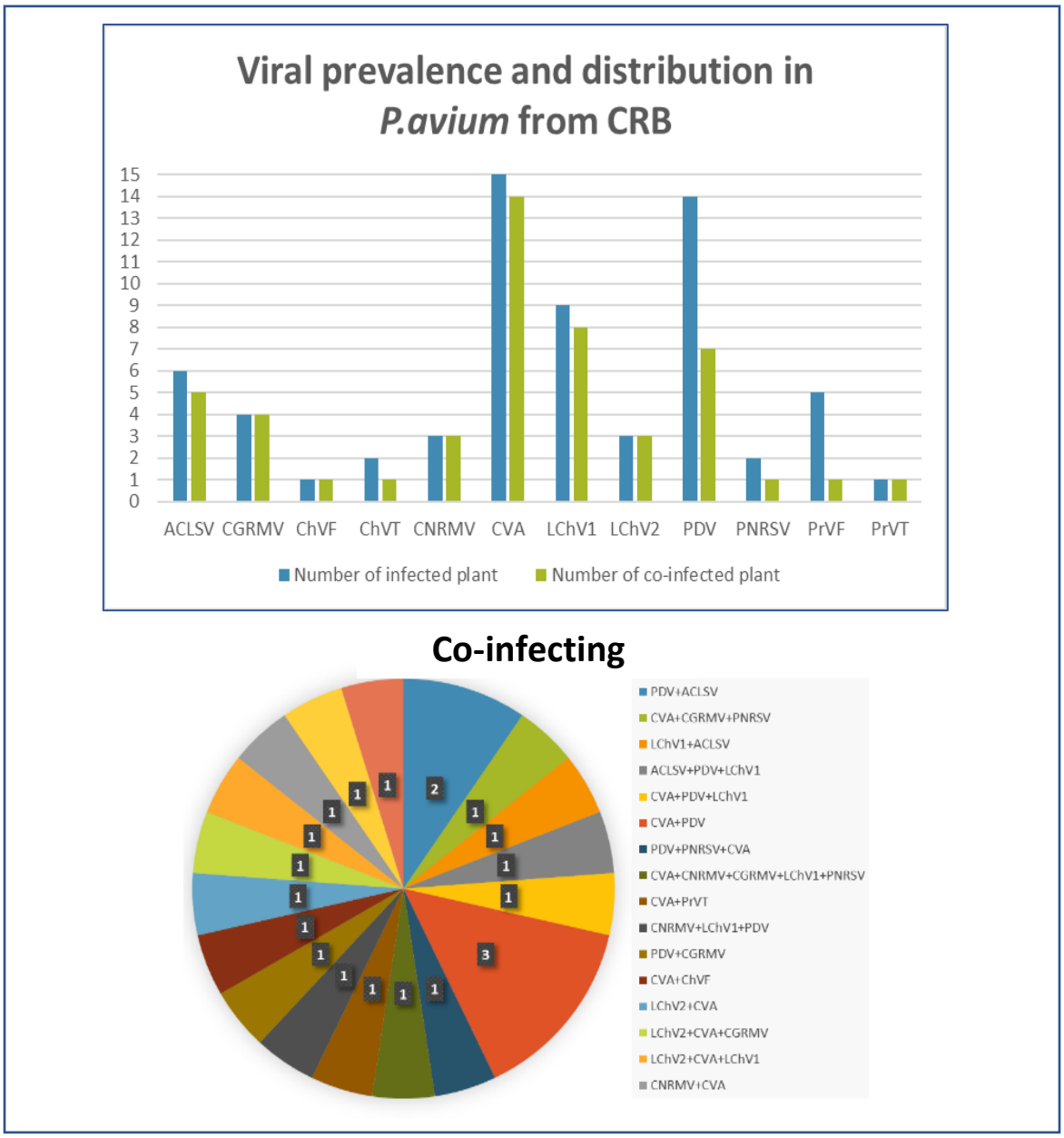


Fig 7. Prevalence of viruses found in sweet cherry in BRC collection and combination of the co-infecting viruses.

3-5 Sour cherry virome

56% (28/50) of the samples were infected by 10 known viruses (CGRMC, ChVF, CNRMV, CVA, LChV1, PDV, PNRSV, PrVF, and ChVT. (Fig. 8). Similar to sweet cherry, CVA was the most prevalent virus in sour cherry. PNRSV and PDV were two other frequent viruses in sour cherry. Unlike sweet cherry, LChV1 was only found in two samples.

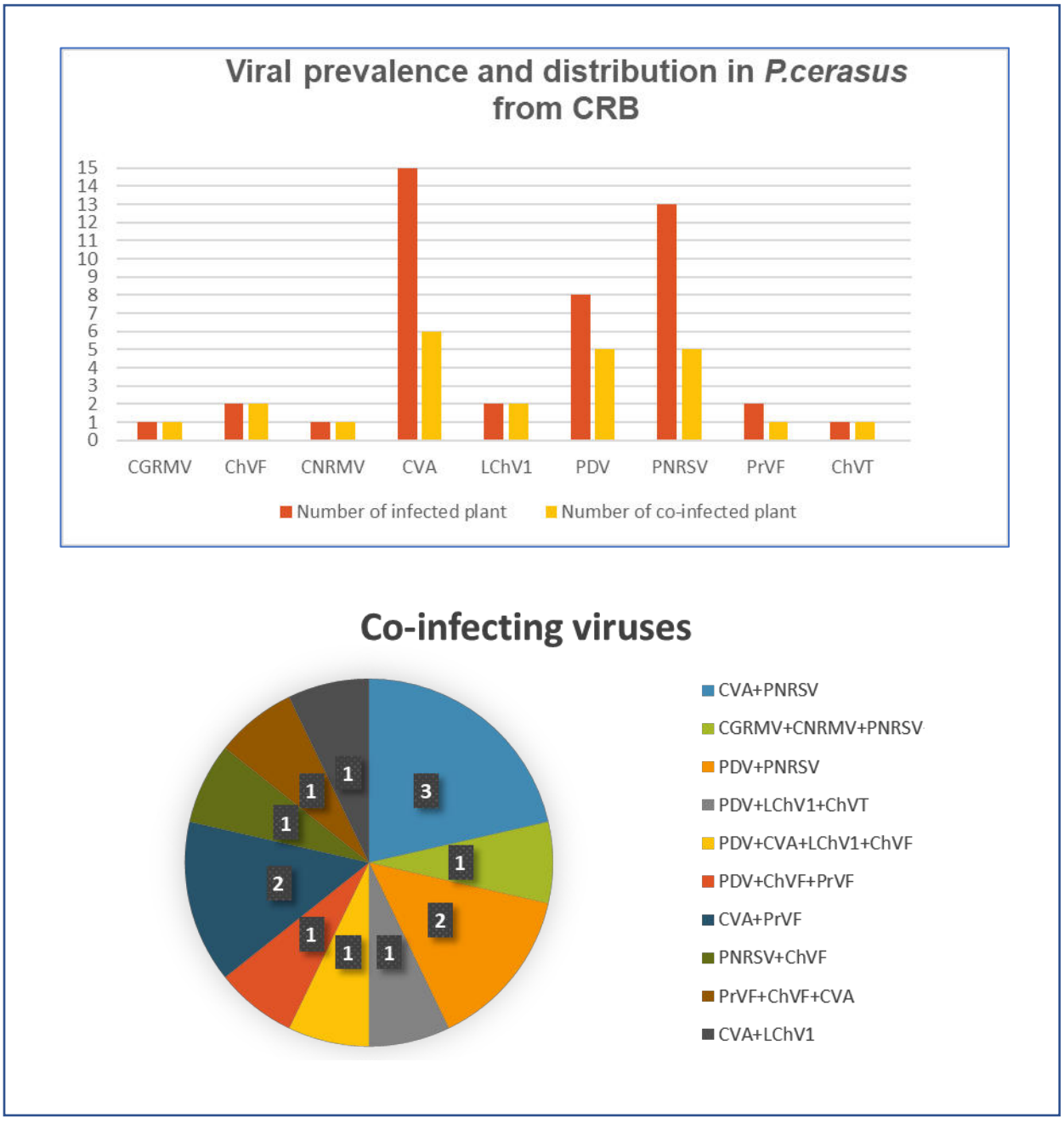


Fig 8. The prevalence of viruses found in sour cherry in BRC collection and combination of the co-infecting viruses.

3-6 Apricot virome

38 % (19/50) of the samples were infected by nine known viruses (CVA, MuVA, APCLSV, ApLV, APV2, LChV1, SLRSV, PNRSV), one novel cheravirus AWPV, described in the chapter III and a known viroid (HSVd) (Fig. 9). The most frequent virus in apricot is PNRSV (16% of the tested trees) followed by LChV1 (8%) and CVA (8%). The co-infection rate was very low in apricot, in comparison to the situation encountered in other *Prunus* species analyzed here and only six trees were coinfected by PNRSV and another virus.

Unlike PNRSV that was found in all five cultivated *Prunus*, ACLSV was not identified in apricot.

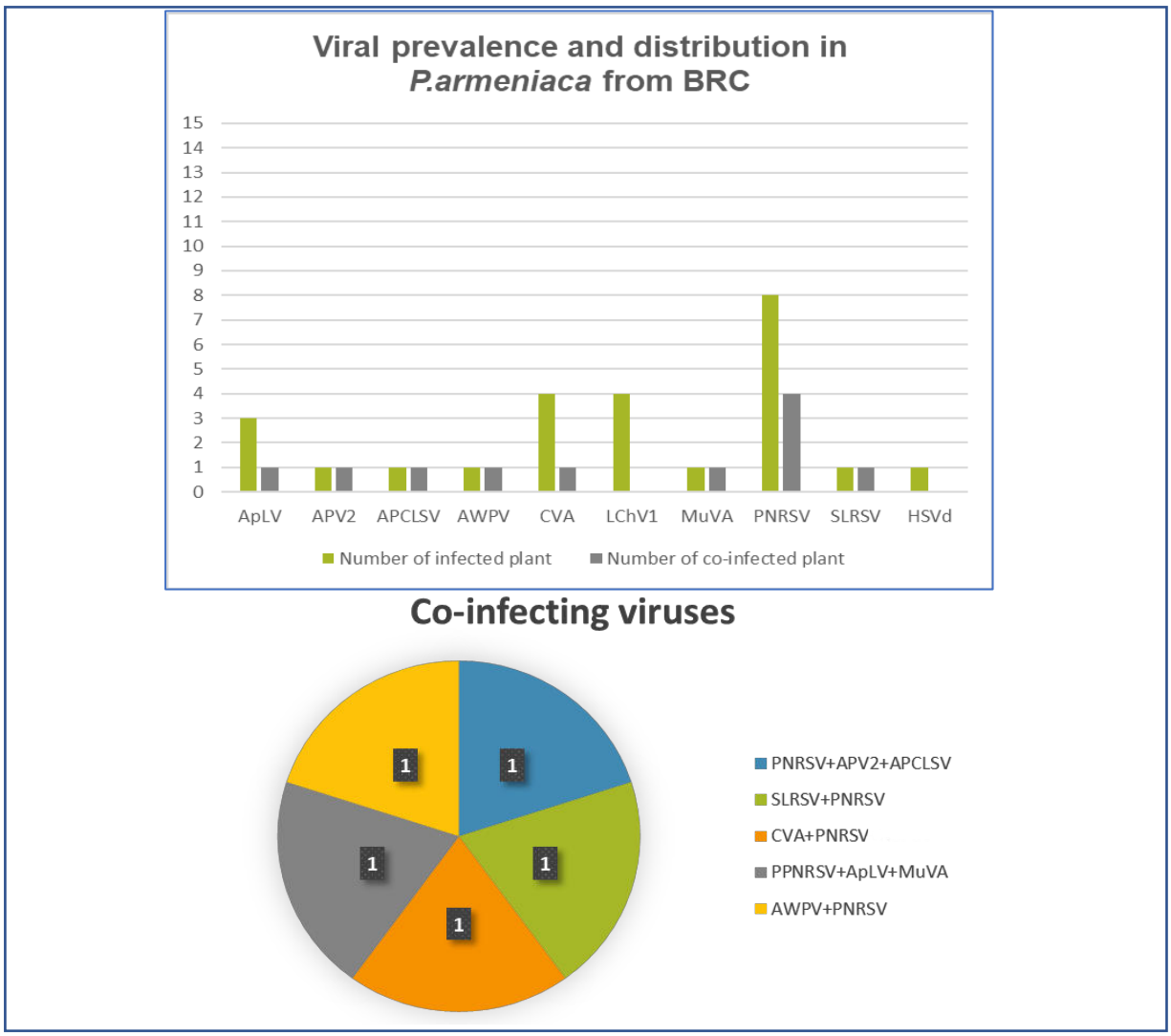


Fig 9. The prevalence of viruses found in apricot in BRC collection and combination of the co-infecting viruses.

3-7 Virome of wild *Prunus* species

The number of viruses identified in wild samples either in germplasms or in the nature was notably less than cultivated species including eight known viruses and three new viruses (Fig. 10). The wild samples collected from Bordeaux in 2020 were tested in pools of 10-15 trees of same species (Table 3). Only PDV and ACLSV could be identified in pooled samples of *P. spinosa*, and only ACLSV in *P. serotina*. In *P. cerasifera*, only PDV and PNRSV could be detected. In contrast, five viruses (CVA, LChV2, ChVF, PDV, and PNRSV) were found infecting *P. serrulata* from the gerplasm collection.

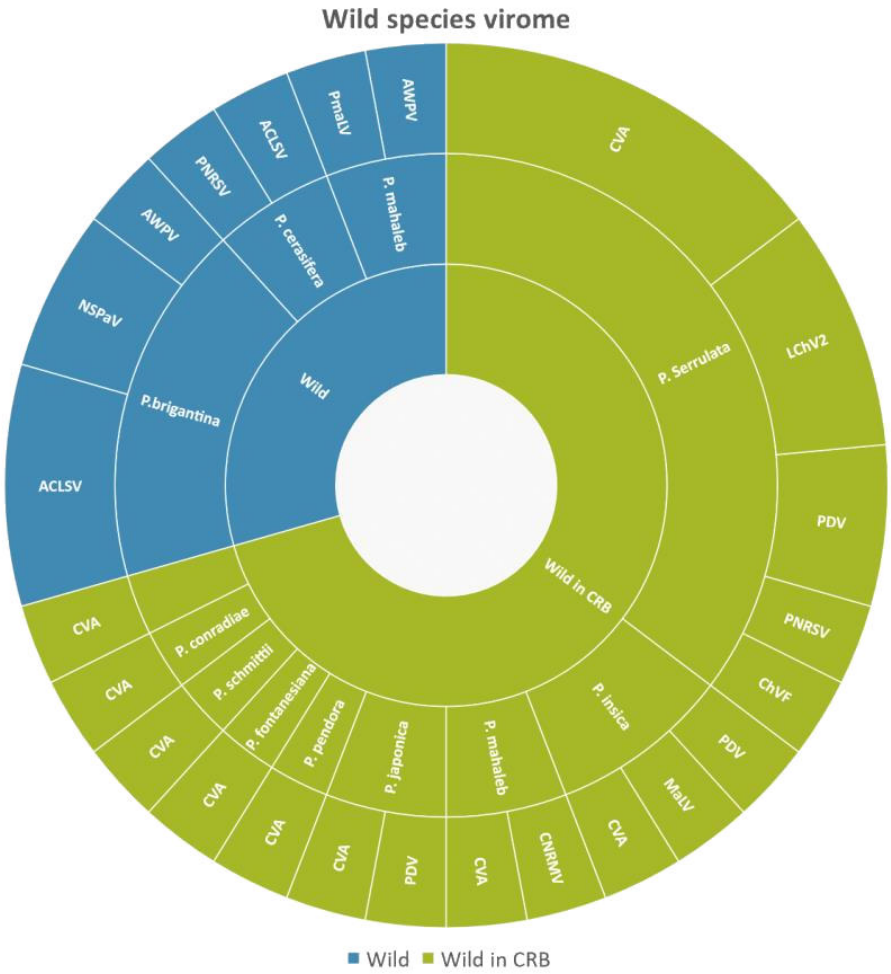


Fig 10. Viruses found in wild species in BRC collection and naturally grown one.

4- Conclusion

Overall, we identified 22 known viruses in addition to two viroids, as well as five novel viruses: a peach-infecting luteovirus (PaLV2), a cheravirus infecting *P. armeniaca* as well as

the wild species *P. brigantina* (AWPV), a luteovirus infecting the wild species *P. mahaleb* (PmaLV), and two other luteoviruses infecting respectively *P. mume* (MaLV) and *P. dulcis* (AlLV1) (see chapters I and II)The virome composition of the five cultivated *Prunus* species analyzed in this study is not very surprising. Apart the novel luteovirus and the novel cheravirus, the virome of peach, sweet and sour cherry, plum, and apricot accessions collected in the BRC is composed of viruses known to infect the hosts in which they have been detected here. Since the approach used for the virome scanning was dsRNA-based viral enrichment, the virome determined comprises the RNA viruses and viroids.

What is more surprising might be the rate of infection, when it comes to the individual trees. This is particularly evident for NSPaV which is detected in 98% of the peach trees sampled and also appears to have a global distribution. Since grafting is the main manner of *Prunus* propagation, the high infection rate of viruses could be due to the rootstock or scion infection. No natural vectors are currently known in luteoviruses infecting *Prunus*, as well as for CVA and LChV1, and ACLSV which were abundantly present in the collection. However,
the possibility of the vector transmission cannot be excluded. The other hypothesis might be the infection of the rootstock used to graft the scion. On the other hand, the high infection rate of PNRSV and PDV can be ascribed to the pollen transmission within the plot.

We performed diversity analysis on some viruses including CVA, PNRSV and LChV1 in plum and ACLSV and NSPaV, PaLV and PaLV2 in peach. Due to the very fragmented assembly of contigs in some dsRNA-HTS data, we did not put any further efforts to perform the diversity analysis for the other viruses and host; but still it can be done through PCR assays.

As metagenomics analyses over the past decades showed the mixed infection of viruses is a rule more than exception (Moreno and Lopez-Moya, 2020), the results of virome study in the present thesis supports that idea. The highest co-infection rate among cultivated species was observed in peach (92%) and the lowest co-infection rate was in apricot (31%).

ANNEX II

VirHunter: A Deep Learning-Based Method for Detection of Novel RNA Viruses in Plant Sequencing Data

frontiers | Frontiers in Bioinformatics

ORIGINAL RESEARCH
published: 13 May 2022
doi: 10.3389/fbinf.2022.867111

VirHunter: A Deep Learning-Based Method for Detection of Novel RNA Viruses in Plant Sequencing Data

Grigorii Sukhorukov[1,2], Maryam Khalili[3], Olivier Gascuel[4], Thierry Candresse[3], Armelle Marais-Colombel[3] and Macha Nikolski[1,2]**

[1]CNRS, IBGC, UMR 5095, Université de Bordeaux, Bordeaux, France, [2]Bordeaux Bioinformatics Center, Université de Bordeaux, Bordeaux, France, [3]Université de Bordeaux, INRAE, UMR BFP, CS20032, CEDEX, Villenave d'Ornon, France, [4]Institut de Systématique, Biodiversité, Evolution (ISYEB - UMR7205, Muséum National d'Histoire Naturelle, CNRS, SU, EPHE, UA), Paris, France

High-throughput sequencing has provided the capacity of broad virus detection for both known and unknown viruses in a variety of hosts and habitats. It has been successfully applied for novel virus discovery in many agricultural crops, leading to the current drive to apply this technology routinely for plant health diagnostics. For this, efficient and precise methods for sequencing-based virus detection and discovery are essential. However, both existing alignment-based methods relying on reference databases and even more recent machine learning approaches are not efficient enough in detecting unknown viruses in RNAseq datasets of plant viromes. We present VirHunter, a deep learning convolutional neural network approach, to detect novel and known viruses in assemblies of sequencing datasets. While our method is generally applicable to a variety of viruses, here, we trained and evaluated it specifically for RNA viruses by reinforcing the coding sequences' content in the training dataset. Trained on the NCBI plant viruses data for three different host species (peach, grapevine, and sugar beet), VirHunter outperformed the state-of-the-art method, DeepVirFinder, for the detection of novel viruses, both in the synthetic leave-out setting and on the 12 newly acquired RNAseq datasets. Compared with the traditional tBLASTx approach, VirHunter has consistently exhibited better results in the majority of leave-out experiments. In conclusion, we have shown that VirHunter can be used to streamline the analyses of plant HTS-acquired viromes and is particularly well suited for the detection of novel viral contigs, in RNAseq datasets.

Keywords: novel virus detection, RNA viruses, plant virome, alignment-free method, deep learning, artificial neural network

OPEN ACCESS

Edited by:
Joao Carlos Setubal, University of São Paulo, Brazil

Reviewed by:
Deyvid Amgarten, Albert Einstein Israelite Hospital, Brazil
Bruno Koshin Vázquez Iha, Universidade Anhanguera de São Paulo, Brazil

****Correspondence:***
Grigorii Sukhorukov
grsukhorukov@gmail.com
Macha Nikolski
macha.nikolski@u-bordeaux.fr

Specialty section:
This article was submitted to Genomic Analysis, a section of the journal Frontiers in Bioinformatics

Received: *31 January 2022*
Accepted: *24 March 2022*
Published: *13 May 2022*

Citation:
Sukhorukov G, Khalili M, Gascuel O, Candresse T, Marais-Colombel A and Nikolski M (2022) VirHunter: A Deep Learning-Based Method for Detection of Novel RNA Viruses in Plant Sequencing Data. Front. Bioinform. 2:867111. doi: 10.3389/fbinf.2022.867111

INTRODUCTION

Study of viromes at an unprecedented scale has been enabled by the adoption of high-throughput sequencing (HTS) technologies and is now frequently undertaken across an increasing range of host species. In particular, sequencing of plant viromes has become quite common, partly due to its relevance to the agricultural sector. The acquired datasets help to elucidate important questions such as virus spread among host reservoirs and effects of agriculture on the ecosystems and their biodiversity as well as the identification of novel viruses in crops and natural environments (Lefeuvre et al., 2019). These developments are fast advancing our knowledge of viral diversity through the

discovery of previously unknown viral species or variants and the identification of new hosts of known viruses (Roossinck et al., 2015; Massart et al., 2017). Following the classification proposed by Stobbe and Roossinck (2014), viruses identified in HTS datasets can be classified into three different groups as follows: 1) viruses that are already known to infect a given host; 2) novel viruses from a known family or known viruses that have not been found previously described to infect a given host; and finally 3) completely novel viruses that share little to no sequence similarity with known viruses already present in the databases.

Using an efficient virus detection method, including for the identification of novel viruses, is essential for efficient disease management. Standard diagnostic tests (ELISA assays and PCR-based assays) depend on specific antibodies or primers and thus require prior knowledge of the virus and of its phylogenetic neighbors. Precise identification of viruses is further complexified by the large diversity encountered in the majority of viral species which is linked to the high mutation rate of these agents. This is particularly true for plant viruses, the majority of which are RNA viruses whose mutation rate is very high (Jenkins et al., 2002). Moreover, the new variants emerging from genomic rearrangements or recombination events can also significantly differ from the parental viruses (Domingo 2010). Also, many of the plant viruses are multihost pathogens, and a single plant can be infected by multiple unrelated viral species (Roossinck, 1997). Such infections by multiple viruses represent an additional challenge for detection since the viral load of different pathogens can be very unequal (Martín and Elena, 2009). Moreover, in most cases, background contamination is currently unavoidable (Kleiner et al., 2015; Maree et al., 2018; Kutnjak et al., 2021). In this context, HTS combined with bioinformatics tools has been shown to be a valuable approach, both for detection of known viruses and for the discovery of novel ones (Maree et al., 2018; Villamor et al., 2019; Mehetre et al., 2021).

Viruses do not have a universal gene marker that could be used for their identification, contrary to the conserved regions of the 16S rRNA and ITS genes, commonly used to classify bacteria and fungi at the genus or species level (Mokili et al., 2012). Moreover, the abundance of viral genomic material in plant sequencing samples can be very low (Massart et al., 2019), due to the dominance of the host material. Hence, specific sample preparation to enrich plant RNA viral-specific sequences is an important step that makes the downstream detection of viruses by bioinformatics methods more reliable. They include approaches providing a high and targeted enrichment of viral sequences, such as the purification of viral double-stranded RNAs (dsRNAs) or that of virion-associated nucleic acids (VANAs) as well as less specific approaches generally affording lower enrichment, such as the sequencing of small interfering RNAs (siRNAs) or inclusion of a ribodepletion step prior to the sequencing of total cellular RNAs (Maree et al., 2018; Kutnjak et al., 2021). As already discussed in a range of reviews, each of these approaches have advantages and weaknesses. In particular, strategies providing high enrichment factors may improve detection sensitivity but often at the cost of introducing biases with the risk of compromising the detection of some particular viruses (Maree et al., 2018; Kutnjak et al., 2021). For example, dsRNA-based approaches are usually poor at detecting DNA viruses, while VANA-based ones may perform poorly for viruses with labile particles.

When interested in known viruses or potentially novel viruses but from a known family, bioinformatics methods that compare the sequenced reads to genomes in public databases are very efficient for virus detection and identification (Stobbe and Roossinck, 2014; Massart et al., 2019). Read-based analysis is thus particularly suited to study viral diversity of sequencing samples in terms of known viral species. Generalistic metagenome analysis tools such as, for example, Kaju (Menzel et al., 2016), Kraken 2 (Wood et al., 2019), and Centrifuge (Kim et al., 2016) show good performance in terms of sensitivity and precision in detection of present known viral species (De Vries et al., 2021).

For the discovery of novel viruses, use of *de novo* assembly to recover novel viral contigs from sequencing data is an essential step in order to overcome the incompleteness of virus reference databases, annotation errors and, importantly, the limited homology between novel viral sequences and reference genomes (Sutton et al., 2019). The assembly step is a staple of short-read sequencing studies, which are still the vast majority today (Maree et al., 2018; Kutnjak et al., 2021). It represents its own challenges, in particular, for very short reads such as those of siRNAs and for viral populations with multiple and microdiverse variants (Warwick-Dugdale et al., 2019), often leading to microdiversity-associated fragmentation and, sometimes, to chimeras in the resulting contigs (Martinez-Hernandez et al., 2017; Roux et al., 2017), which in turn affects the downstream analysis, including estimation of viral diversity and identification of novel viruses (Nayfach et al., 2021). Popular assembler choices are the generalistic de Bruijn graph assembly metaSPAdes (Nurk et al., 2017) and Trinity, for RNAseq (Grabherr et al., 2011).

Following the recent review (Kutnjak et al., 2021), the methods used to analyze assembled contigs can be grouped into three main categories: 1) alignment and mapping-based methods, 2) protein domain searches, and 3) k-mer-based approaches that can either rely on signatures or leverage machine learning. Among this large plethora of tools, alignment-based methods are widely adopted when working with assembled contigs since they provide a longer sequence for homology search against reference genomes using either BLAST (Altschul et al., 1990) and its derivatives or the amino acid alignment of protein-coding genes predicted from the assembled contigs using DIAMOND (Buchfink et al., 2015). Also, focusing the analysis on coding regions is particularly relevant for RNAseq data since the non-coding sequences of viruses are not highly represented in such samples, even if they can be well conserved in certain viral taxa. However, the main drawback of alignment- or mapping-based approaches lies on the fact that they are both computationally intensive and require expertise for filtering and interpreting the results. As for the generalistic k-mer signature approaches, they remain demanding in terms of memory and are best suited for diversity analysis tasks (Kutnjak et al., 2021).

The emergence of machine learning tools for contig-based analysis of virome sequencing data holds much promise to

streamline the discovery of novel viruses in sequencing datasets by both avoiding the time-consuming sequence similarity analyses and modeling even highly divergent sequences. These methods build models based on sequences with known class labels such as "virus" and "host" and learn features that allow them to differentiate between the classes. VirFinder (Ren et al., 2017) and VirSorter2 (Guo et al., 2021) rely on classical machine learning, the former being based on a regularized logistic regression applied to the k-mer frequency matrix extracted from the sequence and the latter on a random forest model built from genomic features. Methods based on deep learning networks have also been proposed for virus detection, such as DeepVirFinder (Ren et al., 2020) and ViraMiner (Tampuu et al., 2019) that both rely on a combination of convolutional neural networks (CNNs) and dense neural networks, and VirNet (Abdelkareem et al., 2018) that relies on a long short-term memory (LSTM) architecture. These three deep learning methods were developed for identification of viral contigs in metagenomic samples and evaluated on bacterial and human metagenomes. However, DeepVirFinder has been recently successfully used in plant-related virome studies (Santos-Medellin et al., 2021).

In this work, we present VirHunter, a deep learning method that uses convolutional neural networks (CNNs), classifies previously assembled contigs to identify potential viral, host, and bacterial (contamination) sequences in RNAseq samples. The hybrid architecture of VirHunter combines a multi-network CNN-based module covering different k-mer sizes with a downstream random forest classifier module. We have trained VirHunter models for three different plant host species: peach, grapevine, and sugar beet. Importantly, we have shown that VirHunter is especially performant for the task of completely novel virus discovery by building 31 leave-out datasets, in which each viral family is excluded from the training dataset, and comparing the results with a standard BLAST-based solution on one side and a state-of-the-art deep learning method, DeepVirFinder, the other side. VirHunter not only systematically outperformed DeepVirFinder in terms of virus detection but also has considerably reduced the False Positive rate. Cross-evaluation has shown that host detection accuracy remained high and decreased slightly when test sequences originated from the plant species were further phylogenetically removed from that used to train the model. We have further evaluated the detection capacity of VirHunter on *in silico* mutated contigs sampled from the NCBI virus database and have shown that it decreased only slightly with a progressively increased mutation rate (e.g., True Positive rate of 0.898 for 20% mutation rate). Moreover, we generated 12 RNAseq datasets for a range of host species and have shown that VirHunter was not only able to uncover the viruses that were previously identified but also to streamline the analyses by considerably reducing the need for manual curation.

MATERIALS AND METHODS

Datasets

We downloaded all complete and incomplete viral sequences from the NCBI virus database for which the host's taxonomic id belongs to *Viridiplantae* on 26/10/2021, which yielded 122,832 sequences. Plant sequences have been downloaded for *Prunus persica* (peach), *Vitis vinifera* (grapevine), *Beta vulgaris* (sugar beet), and *Oryza sativa* (rice) from the NCBI RefSeq genomes database. On one hand, they consist of the latest available assemblies, GCF_000346465.2, GCF_000003745.3, GCF_000511025.2, and GCF_001433935.1 for peach, grapevine, sugar beet, and rice, respectively, and of the coding region sequences (CDSs), on the other hand. In the absence of the plastid sequence in the reference assembly of the sugar beet, we used the separately available sugar beet plastid sequence (NC_059012.1). All complete representative bacterial genomes have been downloaded from the NCBI RefSeq database on 28/10/2021 using the genome_updater.sh script.

To simulate the discovery of completely unknown viruses that do not have expected similarities with the available data, we constructed virus family leave-out datasets by excluding in turns all the sequences of a given plant viral family from the downloaded virus dataset. The NCBI taxonomy contains 45 viral families. We excluded the *Pospiviroidae* and the *Avsunviroidae* families of viroids as they have very small genomes (average length < 1,000). All families with the number of available sequences < 100 were merged in one dataset called *small families*. Finally, all sequences without family labels constituted the *unclassified* dataset. This resulted in 31 leave-out datasets.

Moreover, we generated 12 novel virome-sequencing RNAseq datasets, sampled from peach, grapevine, and sugar beet (see *Sample Preparation and Sequencing*). Description of these datasets and presence of viruses identified by aligning assembled contigs against the NCBI GenBank database (see *Assembly of RNAseq Datasets and Annotation of Viral Contigs*) are listed in the **Supplementary Table S1**.

Sample Preparation and Sequencing

Total RNAs were extracted from three peach leaf samples, three grapevine phloem scrapping samples, and three sugar beet leaf samples using the CTAB method (Chang et al., 1993), the Spectrum™ Plant Total RNA Kit (Sigma-Aldrich, Saint Quentin-Fallavier, France), and the NucleoSpin RNA plant kit (Macherey-Nagel SAS, Hoerdt, France), respectively. RNAseq libraries were prepared either from total RNAs (peach and grapevine samples), messenger RNAs (grapevine samples), or ribodepleted RNAs (sugar beet samples). High-throughput sequencing was performed on an Illumina platform (Hiseq3000 or NovaSeq600) using a paired-end read length of 2 × 150 bp. Accession numbers for each of the three studies (peach, grapevine, and sugar beet) containing raw FASTQ sequencing files are provided in the **Supplementary Table S1**.

Assembly of RNAseq Datasets and Annotation of Viral Contigs

All of the 12 selected plant virome datasets (see *Datasets*) were processed with the QIAGEN CLC Genomics Workbench (v.21.0.5). Briefly, reads were first quality-controlled and trimmed using default parameters and then assembled using

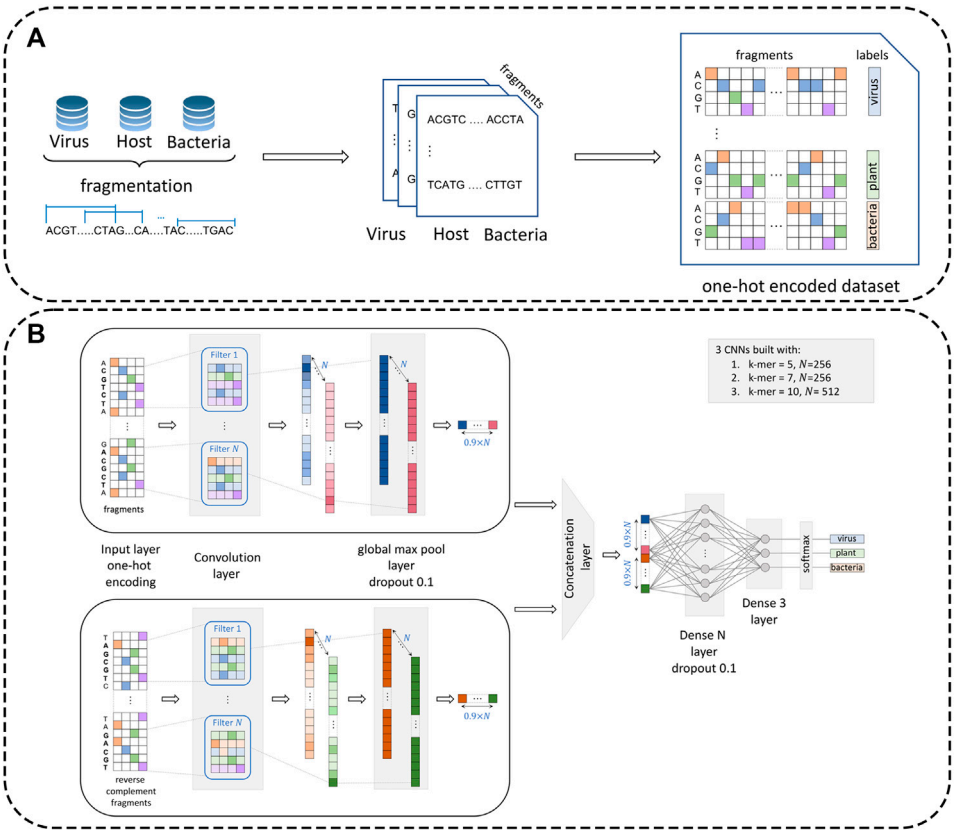


FIGURE 1 | Dataset preprocessing procedure and architecture of the multi-CNN module. Panel **(A)**: Reference datasets (virus, plant, and bacteria) are first fragmented with a pre-defined fragment size (500 and 1000 bp). Each fragment is further one-hot encoded and carries the class label. Panel **(B)**: Three CNN modules are built for k-mers of size $k = 5, 7$, and 10. One-hot encoded genomic fragments of a fixed size are processed by convolutional and global max-pooling layers before being concatenated. A total of two dense layers are followed by the softmax activation function to produce a 3-class classification.

de novo assembly (word size 50, bubble size 300, and minimum contig length 250). To identify viral contigs present in these datasets, we followed a standard three step BLAST-based approach, see e.g., (Candresse et al., 2018). 1) All contigs were aligned using the CLC built in tBLASTx tool against the NCBI nucleotide non-redundant database limited to taxonomic identifiers of viruses. Contigs having significant hits (e-value below the 10^{-20} cut off) were selected. 2) Contigs were further filtered by aligning them using BLASTn and BLASTx with default parameters against the whole GenBank non-redundant nucleotide and protein databases, respectively, and keeping contigs for which the best hits correspond to plant viruses for both BLASTn and BLASTx. Additional manual expert curation allowed to discard contigs with incoherencies between the two alignment results. 3) Finally, all reads passing quality control were mapped against the plant viral contigs, resulting from step 2 using the CLC built-in mapping utility with default parameters with high stringency (90% identity of 90% of read's length). Only contigs with length >750 nucleotides and having sufficient read coverage (expert curation) were retained.

Annotation results together with the counts of thus identified viral contigs are listed in the **Supplementary Table S1**.

Data Preprocessing

To prepare the data for processing by the neural network module, datasets were preprocessed by creating representative one-hot encoded fragments (see **Figure 1**). Specifically, let us denote the virus dataset by V, the plant dataset by H (for "host")—composed of the full assembly G, the coding sequences C, the chloroplast sequence L, and the bacterial dataset by B. Given a fragment size n of 500 and 1,000 nucleotides, V was split into fragments of size n with a sliding window with an increment of $n/2$. Sequences shorter than n nucleotides and longer than $0.95 \times n$ were padded to n bp length with gaps (those shorter than $0.95 \times n$ are discarded), together yielding N viral fragments. Same number N of fragments of size n was randomly sampled from B. As for the plant, G was split into $0.6 \times N$ fragments using a sliding window with an increment of size n, C was split into $0.3 \times N$ fragments with a sliding window with increment of $n/2$, and finally $0.1 \times N$ fragments were sampled randomly from L.

Including plastids in relatively high proportion into the plant dataset H was important to avoid the potential incorrect assignment of contigs originating from plastids to B, given the phylogenetic proximity of plastids and bacteria (McFadden, 2001). Moreover, there are RNA viruses that are known to be replicated in tight association with plastids (mostly chloroplasts) -

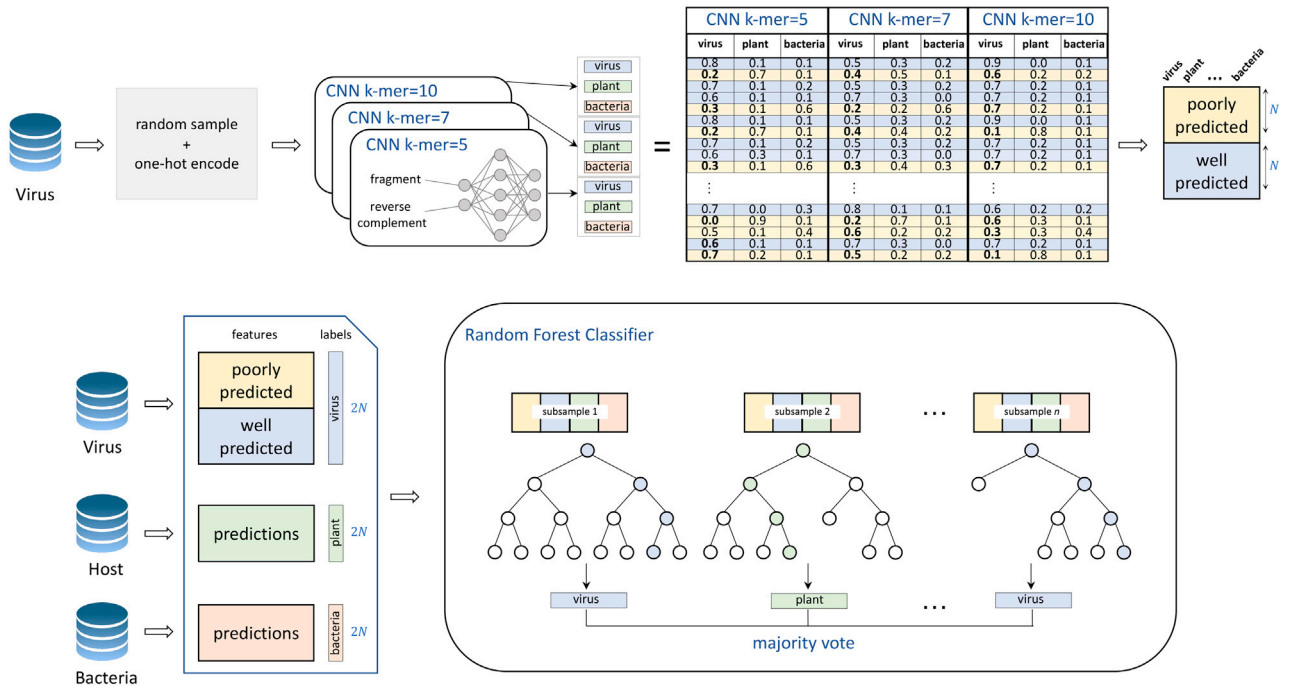


FIGURE 2 | Training of the VirHunter's machine learning module. The individual network predictions are subsetted to contain an equal number of both poorly predicted (prediction value for viral class < 0.8) and well-predicted (prediction value ≥ 0.8) viral fragments (with the goal to overselect poor predictions relative to their overall frequency in order to drive the model to recognize even completely novel viruses). The random forest classifier uses these subsetted predictions for its training.

see e.g., (Budziszewska and Obrępalska-Stęplowska, 2018; Delgado et al., 2019). Enriching for CDS sequences was necessary since the envisioned application of VirHunter is for RNAseq virome datasets. Five compositions of *G*/*C*/*L* proportions of *H* were tested (100/0/0, 90/0/10, 60/30/10, 50/40/10, and 45/45/10, data not shown) and the best was retained.

Fragments were further transformed from length *n* ACGT-character strings to $n \times 4$ one-hot encoded arrays, in which an A is encoded by [1, 0, 0, 0], a C is encoded by [0, 1, 0, 0] *etc*., while gaps are encoded by [0, 0, 0, 0]. Moreover, the encoded dataset is augmented by adding the reverse complement of each original fragment. Indeed, it has been shown by Shrikumar et al. (2017) that CNN models in genomics require the reverse-complement data augmentation combined with parameter sharing between the forward- and reverse-complement representations of the model. Class labels *V*, *H*, or *B* are assigned to each fragment according to its provenance.

VirHunter Architecture

VirHunter architecture was defined with two main components the first component is a multi-path neural network shown in **Figure 1**, and the second component is a machine learning classification module shown in **Figure 2**.

1. Neural network component. The neural network module follows a k-mer-based approach. To alleviate a potential difficulty related to the choice of *k*, VirHunter implements a multi-model solution for $k = 5$, 7, and 10 (see **Figure 1**), with three independent CNN models having the same architecture. These values of *k* were chosen based on the accuracy of the individual CNN networks in the family leave-out experiment (see **Supplemental Figure S1**). The genomic DNA sequence and its reverse complement for each n-size fragment are transformed from nucleotides (in ACGTN alphabet) to an $n \times 4$ one-hot encoded array as presented in *Data Preprocessing*. A convolution layer with leaky rectified linear unit activation function ($a = 0.1$) and global max-pool and dropout layers are then applied independently to the forward fragments and their paired reverse-complement versions. The use of dropout layers was introduced to alleviate the issue of overfitting. Models with $k = 5, 7$ have the convolution layer with 256 filters, while the model for $k = 10$ has 512 filters. The two resulting vectors for the forward- and reverse-complement fragments are then concatenated. Finally, two dense layers are applied. The first dense layer has the number of units equal to 256 for the paths with $k = 5, 7$ and 512 for the path with $k = 10$. It employs a rectified linear unit activation function. The second dense layer has three units and uses the softmax activation function to enable three-class classification.

2. Random Forest component. The second module of the VirHunter implements a random forest classifier (see **Figure 2**) with the goal to aggregate the predictions from three neural networks. The classifier receives nine real-valued predictions from the multi-network module (three per network) and outputs one of the three classes using the majority vote implementation of random forest. The random forest classifier was chosen over other approaches such as linear regression and simple voting, based on performance (data not shown).

Training

The neural network and machine learning modules were trained separately for each of the three plant host species (peach, grapevine, and sugar beet) and for fragment sizes n of 500 and 1,000.

The training dataset for the CNN module was built as presented in *Data Preprocessing*. Training batches with size 512 were prepared in a balanced manner across the three classes (virus, plant, and bacteria) from the training dataset and are split into training and validation with the ratio of 9:1. Each of the three individual networks was trained for 10 epochs, followed by 1 epoch of training on the validation set to take into account all the data.

For training and testing the machine learning components, predictions for the three trained networks were obtained on $100,000$ randomly selected fragments of size n from each V and B. Likewise, $100,000$ fragments of size n were randomly sampled from H, following the ratio described in *Data Preprocessing*. Predictions for random viral fragments were further subsetted in the following manner. We split the test dataset viral fragments into those having good quality predictions (prediction value for viral class ≥ 0.8) and low-quality predictions (prediction value < 0.8) and maintained $10,000$ randomly selected fragments from each category, yielding $20,000$ predictions. These $20,000$ predictions were further selected for plant host H and bacterial B fragments. The resulting dataset with three predictions for each of $60,000$ fragments was further split in train and test datasets with 2:1 ratio, and the machine learning module was trained with parameters max_depth = 5, n_estimators = 10, max_features = 1, and max_samples = 0.2.

We verified that overfitting was successfully circumvented by the individual CNN networks that compose the neural network component of our model by comparing the accuracy on validation and test datasets obtained by these individual networks trained on families in the leave-out experiment for peach (see **Supplementary Table S9**). No significant difference was observed.

Contig Classification

VirHunter trained on fragments with $n = 500$ was used to classify contigs with length $750 < l < 1500$, while VirHunter trained on fragments with $n = 1000$ was used to classify contigs with $1500 < l$. Indeed, an ORF of 500 nucleotides corresponds to an 18 kDa protein, this size covering the vast majority of viral polymerases, movement proteins, and capsid proteins for plant viruses. Contigs with $l < 750$ were considered as very small for prediction by the smaller of the two models and were discarded.

Each fragment of an input contig was preprocessed following the procedure presented in *Data Preprocessing*. Predictions were produced by the neural network module for each of these one-hot encoded fragments, yielding three probabilities of belonging to a specific class (V, H, B). These class probabilities were further processed by the random forest component, resulting in a unique class label for each of the fragments. Finally, given class labels for each of the fragments of the input contig, a vote was applied to decide to which class belongs the whole contig, viral if the number of viral (V) fragments is greater than those from H and from B, host if the number of host (H) fragments is greater than those from V and from B, and bacterial otherwise.

RESULTS

VirHunter Outperforms State-of-the-Art Tools on Family Leave-Out Datasets

VirHunter was trained on GPU (Nvidia Tesla T4) with $n = 1000$ for 31 family leave-out datasets and three different plant datasets (peach, grapevine, and sugar beet), resulting in 63 leave-out models. The test datasets were prepared by random sampling of 30,000 fragments with $n = 1000$ from the corresponding left-aside families of viral sequences, bacteria, and plant.

Classification results for the viral fragments by VirHunter in this family leave-out experiment are shown in **Figure 3** and in **Supplementary Tables S2, S3**. We compared the capacity of VirHunter to detect novel viruses in the family leave-out setting with the BLAST-based approach on one hand and two state-of-the-art machine learning methods, DeepVirFinder and VirSorter2, on the other hand as also shown in **Figure 3**. Briefly, each test dataset was aligned using tBLASTx (v2.12.0), preserving one best hit with parameters -max_target_seqs 1 -max_hsps 1, against the respective virus database with the leave-out family removed, and filtered by e-value $< 10^{-10}$, percent identity > 0.5, and alignment length > 50 amino acids (see results in **Supplementary Table S4**) in order to emulate the annotation workflow without manual inspection; DeepVirFinder was trained on the same 31 leave-out datasets but excluding bacterial fragments from the training dataset since this method provides the possibility to have only two class labels and using the recommended parameters (Ren et al., 2020) on 10 CPUs Intel Xeon CPU E5-2630 v4 (see results in **Supplementary Table S5**); VirSorter2 was evaluated on each test dataset using pretrained models provided by authors (see results in **Supplementary Table S6**).

Variability of correct classification was observed for viral fragments of different left-out families for all three methods as shown in **Figure 3** (see for detailed results in **Supplementary Tables S2–S4**). We have split the families into three groups according to the lowest True Positive (TP) rate of VirHunter across the three plant host species: 21 "easy to classify" (TP rate > 0.7), 7 "moderately difficult to classify" (TP rate between 0.5 and 0.7), and 3 "difficult to classify" (TP rate < 0.5). VirHunter almost systematically outperformed DeepVirFinder in terms of TPs (virus fragments from the leave-out family classified as being viral). In total, there are four exceptions, namely, *Reoviridae*, *Mayoviridae*, *Phycodnaviridae*, and *small families*, out of which *Reoviridae* presented a considerable performance gap. After inspection, it appeared that VirHunter's false negatives for these four families mostly corresponded to viral fragments being classified as bacteria. This is possibly due to the fact that *Mayoviridae* are bacteriophages, *Reoviridae* concern a very wide range of

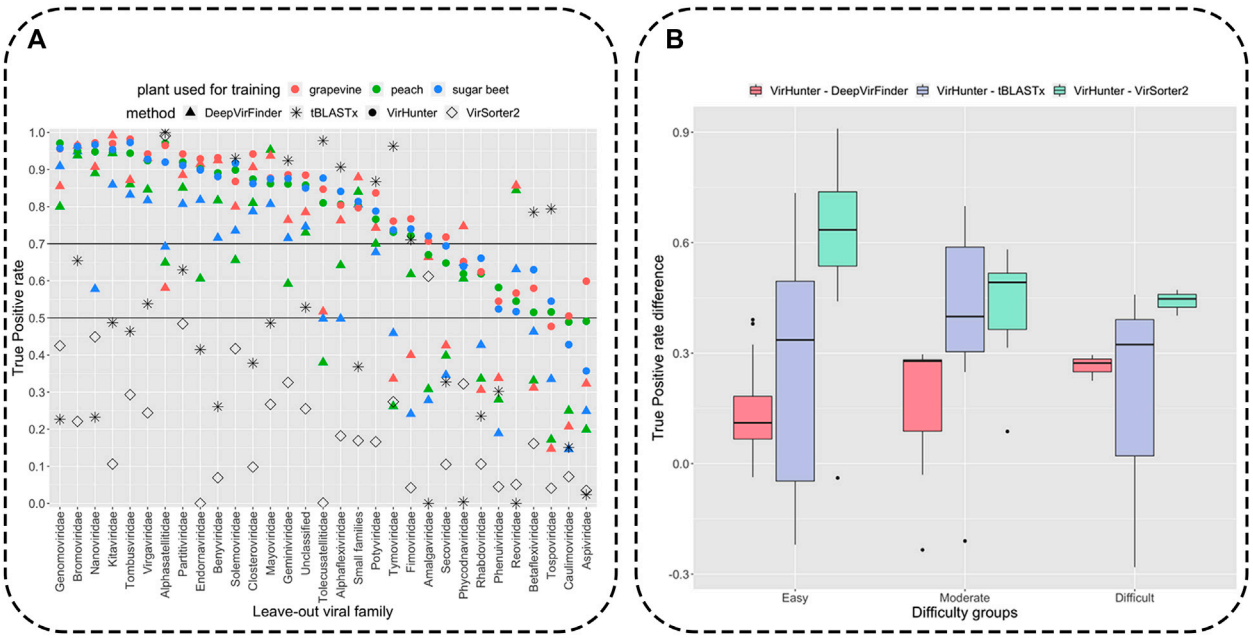


FIGURE 3 | Detection of novel viral fragments in the family leave-out setup. Panel **(A)**: Results for the percent of correctly classified fragments (out of 10,000) with length $n = 1000$ from the corresponding left-aside families. VirHunter results are depicted by circles, tBLASTx by stars, DeepVirFinder by triangles, and VirSorter2 by diamonds. Black lines represent thresholds separating families into three difficulty groups for VirHunter as follows: easy to classify (minimum TP rate across the three plants > 0.7), difficult to classify (minimum TP rate < 0.5), and moderately difficult to classify (minimum TP rate between 0.5 and 0.7). Panel **(B)**: Differences in the True Positive rate between VirHunter, DeepVirFinder (red), tBLASTx (blue), and VirSorter2 (green).

hosts and present characteristics of bacteriophages [likely evolutionary relationship to the *Cystovirus* family of bacteriophage (Guglielmi et al., 2006)], while the *small families* contain a wide variety of viruses, and bacteriophages are one among them (*Mitoviridae*). This is to be counterbalanced by the fact that being trained only on plant and virus sequences due to the 2-class approach, DeepVirFinder systematically erroneously considers the majority of bacterial fragments as being viral (see **Supplementary Table S4**). As for the *Phycodnaviridae* family, it contains dsDNA viruses, which could potentially have contributed to the poorer performance of VirHunter relatively to DeepVirFinder for two of the host species. Altogether, VirHunter has shown consistently better capacity to detect novel viruses than DeepVirFinder.

Of note is also the difference in time requirement for training the VirHunter and DeepVirFinder models. On average, training a full model for one leave-out family for one plant host required 11 h for VirHunter (three CNNs, each for both fragment sizes 500 and 1000 – 6 CNNs in total—and the random forest) and 72 h for DeepVirFinder (four CNNs for fragment sizes 150, 300, 500, and 1000).

Compared to both VirHunter and DeepVirFinder, VirSorter2 has shown poorer performance in the family leave-out setup on all the families except two. Indeed, the TP rate was below 0.5 threshold for all families except for the *Amalgaviridiae* and the *Alphasatellitidae*. For the former, VirSorter outperformed DeepVirFinder, while showing poorer results than VirHunter, while for latter it was the best performing method together with tBLASTx (see Panel A of **Figure 3**).

As shown in **Figure 3**, despite the reasonably permissive filtering criteria, tBLASTx shows best results comparable with VirHunter and for certain families exhibits particularly poor performance relative to the two machine learning methods. For the "easy to classify" families, the difference was mostly in favor of VirHunter, sometimes drastically (see for example, *Nanoviridae* and *Genomoviridae* in Panel A and the boxplot in Panel B). In seven cases, tBLASTx outperformed VirHunter, but this difference was mostly marginal (5.8% difference in TP rate on average), the outlier being *Tolecusatellitidae* and *Tymoviridae*, where the gain in favor of tBLASTx was the strongest. For the "moderately difficult to classify" families, VirHunter had a higher TP rate than tBLASTx in all cases. For the three "difficult to classify" families, even if VirHunter's performance was globally low, it still outperformed tBLASTx, with the notable exception of *Tospoviridae*. Altogether, VirHunter has shown consistently better results than that of tBLASTx, for which the TP rate was frequently below the threshold 0.5 (16 families out of 31).

As for the capacity to correctly classify bacterial fragments, VirHunter has shown a systematically high TP rate, ranging from 0.958 to 0.983, across all the leave-out experiments. As for plant fragments, the TP rate was also satisfactory, sugar beet TP from 0.950 to 0.961, grapevine TP from 0.983 to 0.991, and peach TP

TABLE 1 | VirHunter results for prediction of fragments from different plants. Classification results for three plant-specific models of 10,000 fragments for length 1000 randomly drawn from three plants' reference genomes, from all viral sequences and bacteria are shown. In bold are predictions for the expected class.

Plant used for training	Plant used for testing	Predicted label		
		Plant	Virus	Bacteria
Peach	Peach	**0.988**	0.007	0.006
	Grapevine	**0.892**	0.064	0.044
	Sugar beet	**0.804**	0.113	0.083
	Virus	0.002	**0.996**	0.002
	Bacteria	0.005	0.017	**0.978**
Grapevine	Peach	**0.845**	0.106	0.005
	Grapevine	**0.986**	0.011	0.004
	Sugar beet	**0.78**	0.148	0.072
	Virus	0.002	**0.997**	0.002
	Bacteria	0.007	0.021	**0.973**
Sugar beet	Peach	**0.824**	0.132	0.045
	Grapevine	**0.878**	0.087	0.035
	Sugar beet	**0.956**	0.018	0.026
	Virus	0.002	**0.996**	0.002
	Bacteria	0.012	0.019	**0.969**

from 0.983 to 0.989 (see columns "Bacteria" and "Plant" in **Supplementary Table S2**).

Plant Fragments Are Accurately Classified When VirHunter Is Trained on Phylogenetically Close Plant Species

VirHunter was trained independently with $n = 1000$ for three selected plants (peach, grapevine, and sugar beet) and all the downloaded viruses and bacteria, generated as described in *Data Preprocessing*, yielding three models.

We cross-evaluated VirHunter's ability to correctly predict fragments from the plant absent in the training by sampling from the three studied plants, and 10,000 random fragments with $n = 1000$ were selected. Those three plant test datasets were supplemented with two datasets with $n = 1000$, sampled randomly from all viral sequences and from bacteria, respectively.

As previously described (see *VirHunter Outperforms State-of-the art Tools on Family Leave-out Datasets*), plant fragments, coming from the same plant that the models were trained on, are consistently well classified for all the three models with the TP rate ranging between 0.95 ("sugar beet" model tested on random fragments from the sugar beet genome) and 0.99 (the "peach" model tested on random fragments from the peach genome) as shown in **Table 1**. When the plant host species used for training the model is reasonably phylogenetically close to the one of the test datasets, the impact on the TP rate is not very important. For example, the "peach" model tested on random fragments from the grapevine genome still produces the TP rate of 0.9, and the "grapevine" model tested on peach fragments gives the TP rate of 0.836. However, both these models generate a lower TP rate when tested on random fragments from the more phylogenetically distant sugar beet fragments, 0.827 and 0.781, for the "peach" and "grapevine" models, respectively. Similarly, the "sugar beet" model performs less well for both peach and grapevine random fragments, with TP rates of 0.854 and 0.887, respectively.

The three plants used for training models are phylogenetically distant from one another as they belong to different families, sugar beet belongs to the *Amaranthaceae* family, grapevine belongs to the *Vitaceae* family, and peach to the *Rosaceae* family; all the three are *eudicots*. Out of these three plants, sugar beet is the outlier. Peach and grapevine belong to the *Rosids* higher clade, while sugar beet belongs to the *Caryophyllids* higher clade. Given the phylogenetic distance, the lower bound of 0.78 for the true positive rate between these three plants is reasonable.

To evaluate how strongly the performance would be affected if the host of RNAseq dataset was to be from an even further phylogenetically removed plant (belonging to the *monocots*), we trained a model on the rice (*Oryza sativa*) dataset that belongs to *monocots* higher clade. As shown in the **Supplementary Table S7**, the performance drop was coherent with the increase of the phylogenetic distance (TP rate was 0.766, 0.759, and 0.702 for fragments from peach, grapevine, and sugar beet, respectively); however, the recall remained high for both viral and bacterial fragments. These results highlight that when the host of the RNAseq dataset is phylogenetically highly divergent from any of the plants used to train the available models, a new model for a phylogenetically closer plant has to be trained.

VirHunter Enables Classification of Long Mutated Viral Fragments

To evaluate the potential quality of VirHunter's predictions on contigs' classification, we randomly sampled 10,000 long fragments with $n \in [1500, 2000, 2500, 3000, 4500, 6000]$ from the whole virus dataset V. Furthermore, to better emulate contigs resulting from assembly of sequencing reads, we applied a point mutation rate $m \in [0, 0.05, 0.1, 0.15, 0.2]$ to these long fragments. Classification of the resulting mutated long fragments was performed using models trained for the three plants as described in *VirHunter Enables Classification of Long-Mutated Viral Fragments* and following the procedure for contig classification described in *Contig Classification*.

We observed that VirHunter generated highly accurate predictions for long viral fragments with 0 mutations and that across different fragment sizes (column "Mutation rate" 0 in **Supplementary Table S5**). The TP rate slowly decreased with the increase of the mutation rate: for example, the average TP rate across different fragment sizes with the mutation rate 0.2 was 0.885 for the "peach" model, 0.924 for the "grapevine" model, and 0.885 for the "sugar beet" model. Moreover, these results were consistent between the three plant host species used to build the models: the "peach" model's TP rate was 0.944 in average across different fragment lengths and mutation rates, the "grapevine" models' average TP rate was 0.960, and the "sugar beet" model's average TP rate was 0.936.

TABLE 2 | Performance of VirHunter, DeepVirFinder, and VirSorter2 on 12 RNAseq virome datasets. For each of the 12 datasets shown are the number of contigs that were annotated as being viral by experts and the number of contigs in the initial assembly with length >750. Columns "VirHunter," "DeepVirFinder," and "VirSorter2" show predictions run on these contigs by each method. Columns "# detected" show the total number of contigs detected as being viral by each of the two methods, and columns "detected ∩ annotated" indicates how many of these were previously identified by the curators. Finally, the "tBLASTx e-value < 10^{-10}" column indicates how many of "# detected" contigs align against viruses for VirHunter.

Dataset ID and plant origin		# Contig >750	# Contig annotated as viral	VirHunter			DeepVirFinder		VirSorter2	
				# detected	Detected ⋂ annotated	tBLASTx hits (e-val < 10^{-10})	# detected	Detected ⋂ annotated	# detected	Detected ⋂ annotated
P1	Peach	1,009	2	35	2	14	45	2	10	1
P2	Peach	415	2	19	2	10	32	2	8	1
P3	Peach	685	2	23	2	10	49	2	7	1
G1	Grapevine	9,154	10	153	10	47	133	6	52	4
G2	Grapevine	17,024	10	178	10	40	131	9	117	6
G3	Grapevine	18,750	20	208	18	59	137	17	142	11
G4	Grapevine	4,332	15	95	14	32	81	11	24	4
G5	Grapevine	19,395	25	262	23	73	302	23	144	8
G6	Grapevine	2,932	15	70	14	30	86	13	26	12
S1	Sugar beet	6,082	11	236	10	48	335	11	28	6
S2	Sugar beet	8,902	16	277	16	49	419	16	37	7
S3	Sugar beet	6,912	11	203	11	51	307	11	21	4

VirHunter Uncovers Expected Novel and Known Viral Contigs in Virome

The capacity of VirHunter to detect novel viral contigs from real RNAseq-sequencing data was evaluated and compared to that of DeepVirFinder and VirSorter2. The 12 virome RNAseq datasets, sampled from peach, grapevine, and sugar beet (see **Supplementary Table S1**) were assembled as described in *Assembly of RNAseq Datasets and Annotation of Viral Contigs*. To imitate the novel virus discovery setting, we excluded from the virus dataset those viral species that were annotated as present in the studied plant viromes, and models for each plant species were trained accordingly for VirHunter and DeepVirFinder. For example, to train the "grapevine" model, all viral species present in samples from grapevine (**Supplementary Table S1** column "Present viruses") were deleted from the virus dataset. The same procedure was carried out for training the "peach" and "sugar beet" models. VirSorter2 pretrained models were used following the recommendations in Guo et al. (2021).

The assembled contigs >750 nt were analyzed by VirHunter, DeepVirFinder, and VirSorter2 (see **Table 2** and **Supplementary Table S8**). Importantly, VirHunter assigned a viral label to a lower number of contigs than DeepVirFinder in eight out of 12 datasets ("Viral contigs #" under VirHunter and DeepVirFinder columns). These are the contigs that have to undergo additional manual expert analysis. To better understand their nature, we aligned the contigs identified by VirHunter to the BLAST NCBI nucleotide database limited to "Viruses" taxonomic id as was performed for *Assembly of RNAseq Datasets and Annotation of Viral Contigs* analysis. Contigs getting at least one alignment with percent identity >0.5, length >50 amino acids, and e-value < $10^{-10}$are reported in the column "tBLASTx hits."

Moreover, for six datasets (P1, P2, P3, G4, S2, and S3) VirHunter and DeepVirFinder have correctly identified contigs that were previously annotated as viral. For four datasets (G1, G2, G3, and G6), VirHunter was able to discover additional 4, 3, 5, and 1 contigs, respectively. However, for two cases (G5 and S1), DeepVirFinder identified one more annotated contig relative to VirHunter. While VirSorter2 exhibited lower overprediction comparted to VirHunter and DeepVirFinder, its ability to correctly identify viral contigs was low, as it detected at best 60% of the expected viral contigs.

Remember that contigs annotated by experts were all removed from the virus dataset used for the training of VirHunter and DeepVirFinder, *V*. Consequently, strictly from the computational point of view, detection of these contigs as being viral can thus be considered as detection of novel viruses for those tools. Simple tBLASTx alignment of these expertly annotated contigs against *V* produced variable percent identity, which was as low as 32.4% for a contig from the G1 grapevine dataset and as high as 99% for a contig from the S1 sugar beet dataset (see **Supplementary Table S1**). According to the classification of Stobbe and Roossinck, (2014), discovery of these viruses could thus be assimilated in our setup with the discovery of "novel viruses from a known family" and potentially of "completely novel viruses."

Moreover, it is possible that at least some potentially novel viruses were missed during expert annotation and that the overprediction in columns "# detected" and "tBLASTx hits" (e-val < 10^{-10}) is lower in reality. Indeed, a large number of unidentified novel viruses have been recently shown to be

present in public RNAseq datasets by Edgar et al. (2021), where the authors have identified 10^5 novel RNA viruses. Finally, of note is the considerable gain of time left for expert curation of contigs by approaches similar to that presented in *Assembly of RNAseq Datasets and Annotation of Viral Contigs*, given the numbers in the "# detected" column, where VirHunter has shown improvement over DeepVirFinder in eight out of 12 datasets.

DISCUSSION

High-throughput sequencing (HTS) is capable of broad virus detection for both known and unknown viruses in a variety of hosts and habitats. It has been successfully applied for novel virus discovery in many agricultural crops, leading to the current drive to apply this technology routinely for plant health diagnostics. For this, efficient and precise methods for HTS-based virus detection and discovery are essential.

RNA viruses are the most abundant pathogens infecting plants. However, RNA plant virus detection using HTS presents a number of challenges due to their genetic diversity, lack of conserved regions across viral species, short genome lengths, high mutation rate, and incomplete knowledge present in reference databases. To address this challenge, we developed a novel deep learning method, VirHunter, to detect novel and known plant viruses in RNAseq datasets.

VirHunter is particularly well-suited for the discovery of novel viruses as it was exemplified on 31 synthetic leave-out family datasets, where VirHunter systematically outperformed DeepVirFinder and VirSorter2, reference machine learning tools for virus detection. When compared with the standard tBLASTx approach, we have shown that for most (21 out of 31) leave-out families, VirHunter obtained a higher TP rate. In six cases, tBLASTx was slightly better (5.8% on average). However, there remained four cases where we have seen a much worse performance in VirHunter results. For these specific families, it can be noted that they are particularly well-suited to the alignment-based virus identification, for example, *Alphasatellitidae* viruses carry high sequence similarity to *Geminiviridae* (which was confirmed by the majority of tBLASTx hits).

We have shown that the 3-class classification design of VirHunter, accounting for possible bacterial contamination, was justified by evaluating how such contaminating contigs would be classified. Not surprisingly, VirHunter efficiently dealt with bacterial contamination, while DeepVirFinder classified bacteria mostly (65%) as viruses, which should have been "plants" if the goal is to identify viruses. We have also demonstrated that VirHunter is also perfectly suited for the detection of known divergent viruses, by evaluating classification accuracy on contigs with progressively increasing the mutation rate.

Note the fact that VirHunter is designed to be trained separately for a specific plant host species. However, classification of plant contigs still remained reasonable (minimum 0.78 TP rate) when we performed a cross-evaluation by classifying sequences coming from three phylogenetically distant plants (peach, grapevine, and sugar beet) by each of the three models. As expected, VirHunter performed better, when the plants it was trained and tested on were phylogenetically closer: grapevine and peach belong to the same *rosids* higher clade resulted in better mutual predictions, while sugar beet as an outgroup belonging to the *caryophyllids* higher clade has shown a relative drop in performance. All these three plants are *eudicots* (Pin 2012). When the model was trained on an even further phylogenetically distant plant, rice that belongs to *monocots* and tested on fragments from peach, grapevine, and sugar beet, the classification accuracy of VirHunter was expectedly lower. Together this implies that to classify contigs from an RNAseq experiment, using a pretrained model trained on the exact same plant species as the host of the experimental dataset is not mandatory, but it is preferable to use one trained on a phylogenetically close plant, ideally from the same family and at least belonging to the same *eudicots/monocots* higher clade. A possible avenue to explore in the future work is the feasibility of transfer learning (Eraslan et al., 2019), to enable fast on-demand retraining for a new plant or building a generalistic plant model.

Finally, we validated VirHunter's capacity to detect novel viruses on 12 newly acquired RNAseq datasets for peach, grapevine, and sugar beet. In these datasets, VirHunter detected at least 90% (73% for DeepVirFinder and 26% for VirSorter2) of all expert-annotated viral contigs, and in seven datasets it was 100%. Another contribution is the low rate of false positives generated by VirHunter, leaving from 19 to 277 contigs depending on the dataset to be inspected by an expert. These results indicate that VirHunter efficiently reduces the number of contigs requiring manual expert curation.

In conclusion, we have shown that VirHunter can be used to streamline the analyses of plant HTS-acquired viromes and is particularly well suited for the detection of novel viral contigs, in RNAseq datasets.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

MN and OG conceptualized the approach. MN and TC designed the study. MN, TC, and AM-C supervised the research. MN and GS contributed to the computational experimental design. GS implemented VirHunter. GS and MK performed genome assembly. TC, AM-C, and MK collected the samples and generated sequencing data. MK and TC performed data annotation. All the authors contributed to writing the manuscript.

FUNDING

This work was supported by the funding from Horizon 2020 Marie Skłodowska-Curie Actions Innovative Training Network (H2020 MSCA-ITN) project "INEXTVIR" (GA 813542).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbinf.2022.867111/full#supplementary-material

REFERENCES

Abdelkareem, A. O., Khalil, M. I., Elaraby, M., Abbas, H., and Elbehery, A. H. A. (2018). "VirNet: Deep Attention Model for Viral Reads Identification," in Procs of the 2018 13th Intl Conf. on Computer Engineering and Systems (ICCES), Cairo, Egypt, 18-19 Dec. 2018, 623–626. doi:10.1109/icces.2018.8639400

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic Local Alignment Search Tool. *J. Mol. Biol.* 215 (3), 403–410. doi:10.1016/S0022-2836(05)80360-2

Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and Sensitive Protein Alignment Using DIAMOND. *Nat. Methods* 12, 59–60. doi:10.1038/nmeth.3176

Budziszewska, M., and Obrępalska-Stęplowska, A. (2018). The Role of the Chloroplast in the Replication of Positive-Sense Single-Stranded Plant RNA Viruses. *Front. Plant Sci.* 9, 1776. doi:10.3389/fpls.2018.01776

Candresse, T., Theil, S., Faure, C., and Marais, A. (2018). Determination of the Complete Genomic Sequence of Grapevine Virus H, a Novel Vitivirus Infecting grapevine. *Arch. Virol.* 163, 277–280. doi:10.1007/s00705-017-3587-7

Chang, S., Puryear, J., and Cairney, J. (1993). A Simple and Efficient Method for Isolating RNA from pine Trees. *Plant Mol. Biol. Rep.* 11, 113–116. doi:10.1007/BF02670468

de Vries, J. J. C., Brown, J. R., Fischer, N., Sidorov, I. A., Morfopoulou, S., Huang, J., et al. (2021). Benchmark of Thirteen Bioinformatic Pipelines for Metagenomic Virus Diagnostics Using Datasets from Clinical Samples. *J. Clin. Virol.* 141, 104908. doi:10.1016/j.jcv.2021.104908

Delgado, S., Navarro, B., Serra, P., Gentit, P., Cambra, M. Á., Chiumenti, M., et al. (2019). How Sequence Variants of a Plastid-Replicating Viroid with One Single Nucleotide Change Initiate Disease in its Natural Host. *RNA Biol.* 16 (7), 906–917. doi:10.1080/15476286.2019.1600396

Domingo, E. (2010). Mechanisms of Viral Emergence. *Vet. Res.* 41, 38. doi:10.1051/vetres/2010010

Edgar, R. C., Taylor, J., Lin, V., Altman, T., Barbera, P., Meleshko, D., et al. (2021). *Petabase-scale Sequence Alignment Catalyses Viral Discovery*. BioRxiv. 2020-08.

Eraslan, G., Avsec, Ž., Gagneur, J., and Theis, F. J. (2019). Deep Learning: New Computational Modelling Techniques for Genomics. *Nat. Rev. Genet.* 20, 389–403. doi:10.1038/s41576-019-0122-6

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length Transcriptome Assembly from RNA-Seq Data without a Reference Genome. *Nat. Biotechnol.* 29 (7), 644–652. doi:10.1038/nbt.1883

Guglielmi, K. M., Johnson, E. M., Stehle, T., and Dermody, T. S. (2006). Attachment and Cell Entry of Mammalian Orthoreovirus. *Curr. Top. Microbiol. Immunol.* 309, 1–38. doi:10.1007/3-540-30773-7_1

Guo, J., Bolduc, B., Zayed, A. A., Varsani, A., Dominguez-Huerta, G., and Delmont, T. O. (2021). VirSorter2: A Multi-Classifier, Expert-Guided Approach to Detect Diverse DNA and RNA Viruses. *Microb.* 9 (1), 1–13. doi:10.1186/s40168-020-00990-y

Jenkins, G. M., Rambaut, A., Pybus, O. G., and Holmes, E. C. (2002). Rates of Molecular Evolution in RNA Viruses: A Quantitative Phylogenetic Analysis. *J. Mol. Evol.* 54, 156–165. doi:10.1007/s00239-001-0064-3

Kim, D., Song, L., Breitwieser, F. P., and Salzberg, S. L. (2016). Centrifuge: Rapid and Sensitive Classification of Metagenomic Sequences. *Gen. Res.* 26 (12), 1721–1729. doi:10.1101/gr.210641.116

Kleiner, M., Hooper, L. V., and Duerkop, B. A. (2015). Evaluation of Methods to Purify Virus-like Particles for Metagenomic Sequencing of Intestinal Viromes. *BMC Genomics* 16 (1), 7–15. doi:10.1186/s12864-014-1207-4

Kutnjak, D., Tamisier, L., Adams, I., Boonham, N., Candresse, T., Chiumenti, M., et al. (2021). A Primer on the Analysis of High-Throughput Sequencing Data for Detection of Plant Viruses. *Microorganisms* 9, 841. doi:10.3390/microorganisms9040841

Lefeuvre, P., Martin, D. P., Elena, S. F., Shepherd, D. N., Roumagnac, P., and Varsani, A. (2019). Evolution and Ecology of Plant Viruses. *Nat. Rev. Microbiol.* 17, 632–644. doi:10.1038/s41579-019-0232-3

Maree, H. J., Fox, A., Al Rwahnih, M., Boonham, N., and Candresse, T. (2018). Application of HTS for Routine Plant Virus Diagnostics: State of the Art and Challenges. *Front. Plant Sci.* 9, 1082. doi:10.3389/fpls.2018.01082

Martín, S., and Elena, S. F. (2009). Application of Game Theory to the Interaction between Plant Viruses during Mixed Infections. *J. Gen. Virol.* 90, 2815–2820. doi:10.1099/vir.0.012351-0

Martinez-Hernandez, F., Fornas, O., Lluesma Gomez, M., Bolduc, B., de la Cruz Peña, M. J., Martínez, J. M., et al. (2017). Single-virus Genomics Reveals Hidden Cosmopolitan and Abundant Viruses. *Nat. Commun.* 8, 15892. doi:10.1038/ncomms15892

Massart, S., Candresse, T., Gil, J., Lacomme, C., Predajna, L., Ravnikar, M., et al. (2017). A Framework for the Evaluation of Biosecurity, Commercial, Regulatory, and Scientific Impacts of Plant Viruses and Viroids Identified by NGS Technologies. *Front. Microbiol.* 8, 45. doi:10.3389/fmicb.2017.00045

Massart, S., Chiumenti, M., De Jonghe, K., Glover, R., Haegeman, A., Koloniuk, I., et al. (2019). Virus Detection by High-Throughput Sequencing of Small RNAs: Large-Scale Performance Testing of Sequence Analysis Strategies. *Phytopathology* 109 (3), 488–497. doi:10.1094/PHYTO-02-18-0067-R

McFadden, G. I. (2001). Primary and Secondary Endosymbiosis and the Origin of Plastids. *J. Phycology* 37 (6), 951–959. doi:10.1046/j.1529-8817.2001.01126.x

Mehetre, G. T., Leo, V. V., Singh, G., Sorokan, A., Maksimov, I., Yadav, M. K., et al. (2021). Current Developments and Challenges in Plant Viral Diagnostics: A Systematic Review. *Viruses* 13, 412. doi:10.3390/v13030412

Menzel, P., Ng, K. L., and Krogh, A. (2016). Fast and Sensitive Taxonomic Classification for Metagenomics With Kaiju. *Nature Communications* 7 (1), 1–9. doi:10.1038/ncomms11257

Mokili, J. L., Rohwer, F., and Dutilh, B. E. (2012). Metagenomics and Future Perspectives in Virus Discovery. *Curr. Opin. Virol.* 2, 63–77. doi:10.1016/j.coviro.2011.12.004

Nayfach, S., Camargo, A. P., Schulz, F., Eloe-Fadrosh, E., Roux, S., and Kyrpides, N. C. (2021). CheckV Assesses the Quality and Completeness of Metagenome-Assembled Viral Genomes. *Nat. Biotechnol.* 39, 578–585. doi:10.1038/s41587-020-00774-7

Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P. A. (2017). metaSPAdes: a New Versatile Metagenomic Assembler. *Genome Res.* 27 (5), 824–834. doi:10.1101/gr.213959.116

Pin, P. A. (2012). *Life Cycle and Flowering Time Control in Beet*. PhD Thesis (Sweden, Umeå: Swedish University of Agricultural Sciences).

Ren, J., Ahlgren, N. A., Lu, Y. Y., Fuhrman, J. A., and Sun, F. (2017). VirFinder: A Novel k-mer Based Tool for Identifying Viral Sequences From Assembled Metagenomic Data. *Microb.* 5 (1), 1–20. doi:10.1186/s40168-017-0283-5

Ren, J., Song, K., DengAhlgren, C. N. A., Ahlgren, N. A., Fuhrman, J. A., Li, Y., et al. (2020). Identifying Viruses from Metagenomic Data Using Deep Learning. *Quant. Biol.* 8, 64–77. doi:10.1007/s40484-019-0187-4

Roossinck, M. J., Martin, D. P., and Roumagnac, P. (2015). Plant Virus Metagenomics: Advances in Virus Discovery. *Phytopathology* 105, 716–727. doi:10.1094/PHYTO-12-14-0356-RVW

Roossinck, M. J. (1997). Mechanisms of Plant Virus Evolution. *Annu. Rev. Phytopathol* 35, 191–209. doi:10.1146/annurev.phyto.35.1.191

Rott, M., Xiang, Y., Boyes, I., Belton, M., Saeed, H., Kesanakurti, P., et al. (2017). Application of Next Generation Sequencing for Diagnostic Testing of Tree Fruit Viruses and Viroids. *Plant Dis.* 101, 1489–1499. doi:10.1094/PDIS-03-17-0306-RE

Roux, S., Emerson, J. B., Eloe-Fadrosh, E. A., and Sullivan, M. B. (2017). Benchmarking Viromics: an In Silico Evaluation of Metagenome-Enabled Estimates of Viral Community Composition and Diversity. *PeerJ* 5, e3817. doi:10.7717/peerj.3817

Santos-Medellin, C., Zinke, L. A., Ter Horst, A. M., Gelardi, D. L., Parikh, S. J., and Emerson, J. B. (2021). Viromes Outperform Total Metagenomes in Revealing the Spatiotemporal Patterns of Agricultural Soil Viral Communities. *The ISME Journ* 15, 1–15. doi:10.1038/s41396-021-00897-y

Shrikumar, A., Greenside, P., and Kundaje, A. (2017). *Reverse-complement Parameter Sharing Improves Deep Learning Models for Genomics*. bioRxiv. 103663.

Stobbe, A. H., and Roossinck, M. J. (2014). Plant Virus Metagenomics: What We Know and Why We Need to Know More. *Front. Plant Sci.* 5, 150. doi:10.3389/fpls.2014.00150

Sutton, T. D. S., Clooney, A. G., Ryan, F. J., Ross, R. P., and Hill, C. (2019). Choice of Assembly Software Has a Critical Impact on Virome Characterisation. *Microbiome* 7 (1), 12–15. doi:10.1186/s40168-019-0626-5

Tampuu, A., Bzhalava, Z., Dillner, J., and Vicente, R. (2019). ViraMiner: Deep Learning on Raw DNA Sequences for Identifying Viral Genomes in Human Samples. *PLoS ONE* 14, e0222271. doi:10.1371/journal.pone.0222271

Villamor, D. E. V., Ho, T., Al Rwahnih, M., Martin, R. R., and Tzanetakis, I. E. (2019). High Throughput Sequencing for Plant Virus Detection and Discovery. *Phytopathology* 109 (5), 716–725. doi:10.1094/PHYTO-07-18-0257-RVW

Warwick-Dugdale, J., Solonenko, N., Moore, K., Chittick, L., Gregory, A. C., Allen, M. J., et al. (2019). Long-read Viral Metagenomics Captures Abundant and Microdiverse Viral Populations and Their Niche-Defining Genomic Islands. *PeerJ* 7, e6800. doi:10.7717/peerj.6800

Wood, D. E., Lu, J., and Langmead, B. (2019). Improved Metagenomic Analysis With Kraken 2. *Gen. Biol.* 20 (1), 1–13. doi:10.1186/s13059-019-1891-0

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Sukhorukov, Khalili, Gascuel, Candresse, Marais-Colombel and Nikolski. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

BIBLIOGRAPHY

Bibliography

Abdelkareem A. O., Khalil M. I., Elaraby M., Abbas H. &. Elbehery A. H. A. (2018). VirNet: Deep attention model for viral reads identification. Pages 623-626, In: *13th International Conference on Computer Engineering and Systems*. DOI: 10.1109/ICCES.2018.8639400.

Adams, M. J., Candresse, T., Hammond, J., Kreuze, J. F., Martelli, G. P., Namba, S., Pearson, M.N., Ryu, K.H., Saldarelli, P. and Yoshikawa, N. (2012). Family Betaflexiviridae. Pages 920-941, In: *Virus taxonomy: 9th report of the International Committee on Taxonomy of Viruses.* Adams, M. J., Lefkowitz, E. J., Carstens, E. B. and King, A. M. Q. (eds.) Elsevier Academic Press, San Diego, CA. DOI: 10.1016/B978-0-12-384684-6.00078-1

Adams, I. A., Fox, A., Boonham, N., Massart, S., & De Jonghe, K. (2018) The impact of high throughput sequencing on plant health diagnostics. *European Journal of Plant Pathology, 152,* 909–919. DOI: 10.1007/s10658-018-1570-0

Adams, I. P., Glover, R. H., Monger, W. A., Mumford, R., Jackeviciene, E., Navalinskiene, M., Samuitiene, M., & Boonham, N. (2009). Next-generation sequencing and metagenomic analysis: a universal diagnostic tool in plant virology. *Molecular Plant Pathology*, *10*, 537–545. DOI: 10.1111/j.1364-3703. 2009.00545.x

Aiewsakun, P., & Katzourakis, A. (2015). Endogenous viruses: Connecting recent and ancient viral evolution. *Virology, 479,* 26-37. DOI: 10.1016/j.virol.2015.02.011

Al Rwahnih, M., Alabi, O. J., Westrick, N. M., & Golino, D. (2018). *Prunus* geminivirus A: a novel Grablovirus infecting *Prunus* spp. *Plant disease*, *102*, 1246-1253. DOI: 10.1094/PDIS-09-17-1486-RE

Al Rwahnih, M., Daubert, S., Golino, D., & Rowhani, A. (2009). Deep sequencing analysis of RNAs from a grapevine showing Syrah decline symptoms reveals a multiple virus

infection that includes a novel virus. *Virology*, *387*, 395–401. DOI: 10.1016/j.virol.2009.02.028

Al Rwahnih, M., Daubert, S., Golino, D., Islas, C., & Rowhani, A. (2015). Comparison of Next-Generation Sequencing Versus Biological Indexing for the Optimal Detection of Viral Pathogens in Grapevine. *Phytopathology*, *105*, 758–763. DOI: 10.1094/phyto-06-14-0165-r

Aparicio, F., Myrta, A., Di Terlizzi, B., &Pallás, V. (1999). Molecular Variability Among Isolates of *Prunus* Necrotic Ringspot Virus from Different *Prunus* spp. *Phytopathology*, *89*, 991-999. DOI: 10.1094/PHYTO.1999.89.11.991.

Anderson, P. K., Cunningham, A. A., Patel, N. G., Morales, F. J., Epstein, P. R., & Daszak, P. (2004). Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends in Ecology & Evolution*, *19*, 535–544. DOI: 10.1016/j.tree.2004.07.021

Atanasov, D. (1933). Plum pox. A new virus disease. Annuaire de L'Universite de Sofia. *Faculte D'Agronomie et de Sylviculture*, *9*, 49-70.

Bag, S., Al Rwahnih, M., Li, A., Gonzalez, A., Rowhani, A., Uyemoto, J. K., & Sudarshana, M. R. (2015). Detection of a new luteovirus in imported nectarine trees: a case study to propose adoption of metagenomics in post-entry quarantine. *Phytopathology*, *105*, 840-846. DOI: 10.1094/PHYTO-09-14-0262-R

Baizan-Edge, A., Cock, P., MacFarlane, S., McGavin, W., Torrance, L., Jones, S. Kodoja. (2019). A workflow for virus detection in plants using k-mer analysis of RNA-sequencing data. *Journal of General Virology*, *100*, 533-542. DOI: 10.1099/jgv.0.001210.

Barath, D., Jaksa-Czotter, N., Varga, T., & Varallyay, E. (2022). Viromes of Hungarian Peach Trees Identified by High-Throughput Sequencing of Small RNAs. *Plants*, *11*, 1591. DOI: 10.3390/plants11121591

Baulcombe D. (2005). RNA silencing. *Trends in Biochemical Sciences, 30*, 290-293. DOI: 10.1016/j.tibs.2005.04.012.

Bejerman, N., Debat, H., & Dietzgen RG. (2020). The Plant Negative-Sense RNA Virosphere: Virus Discovery Through New Eyes. *Frontiers in Microbiology, 11,* 588427. DOI: 10.3389/fmicb.2020.588427.

Bernardo, P., Charles-Dominique, T., Barakat, M., Ortet, P., Fernandez, E., Filloux, D., Hartnady, P., Rebelo, T. A., Cousins, S. R., Mesleard, F., Cohez, D., Yavercovski, N., Varsani, A., Harkins, G. W., Peterschmitt, M., Malmstrom, C. M., Martin, D. P., & Roumagnac, P. (2017). Geometagenomics illuminates the impact of agriculture on the distribution and prevalence of plant viruses at the ecosystem scale. *The ISME Journal*, *12* 173–184. DOI: 10.1038/ismej.2017.155

Bester, R., Malan, S. S., &Maree, H. J. (2020). A Plum Marbling Conundrum: Identification of a New Viroid Associated with Marbling and Corky Flesh in Japanese Plums. *Phytopathology. 110,* 1476-1482. DOI: 10.1094/PHYTO-12-19-0474-R.

Blando, F., & Oomah, B. D. (2019). Sweet and sour cherries: Origin, distribution, nutritional composition and health benefits. *Trends in Food Science & Technology*, *86*, 517-529. DOI: /10.1016/j.tifs.2019.02.052

Bordenstein, S. R., & Theis, K. R. (2015). Host Biology in Light of the Microbiome: Ten Principles of Holobionts and Hologenomes. *PLoS Biology 18*, e1002226. DOI: 10.1371/journal.pbio.1002226.

Bradford Hill, A. (1965). The environment and disease: Association or causation? *Proceedings of the* Royal Society *of Medicine, 58*, 295-300

Brewer, E., Cao, M., Gutierrez, B., Bateman, M., & Li, R. (2020). Discovery and molecular characterization of a novel trichovirus infecting sweet cherry. *Virus genes*, *56*, 380-385. DOI: 10.1007/s11262-020-01743-7

Bronzato Badial, A., Sherman, D., Stone, A., Gopakumar, A., Wilson, V., Schneider, W., & King, J. (2018). Nanopore Sequencing as a Surveillance Tool for Plant Pathogens in Plant and Insect Tissues. *Plant Disease, 102,* 1648-1652. DOI: 10.1094/PDIS-04-17-0488-RE.

Bujarski, J., Figlerowicz, M., Gallitelli, D., Roossinck, M. J., & Scott, S. W. (2019). ICTV Virus Taxonomy Profile: *Bromoviridae*. *Journal of General Virology, 100,* 1206-1207. DOI: 10.1099/jgv.0.001282.

Çağlayan, K., Roumi, V., Gazel, M., Elçi, E., Acioğlu, M., Mavric Plesko, I., Reynard, J. S., Maclot, F., & Massart, S. (2019). Identification and characterization of a novel Robigovirus species from sweet cherry in turkey. *Pathogens, 8,* 57. DOI: 10.3390/pathogens8020057

Cambra, M., Capote, N., Myrta, A., & Llácer, G. (2006). Plum pox virus and the estimated costs associated with sharka disease. *EPPO bulletin, 36,* 202-204. DOI: 10.1111/j.1365-2338.2006.01027.x

Candresse, T., Delbos, R. P., Le Gall, O., Dunez, J., & Desvignes, J. C. (1998). Characterization of Stocky prune virus, a new Nepovirus detected in French prunes. *Acta Horticulturae, 472,* 175-182. DOI: 10.17660/ActaHortic.1998.472.18

Candresse, T., Filloux, D., Muhire, B., Julian, C., Galzi, S., Fort, G., Bernardo, P., Daugrois, J. H., Fernandez, E., Martin, D. P., Varsani, A., & Roumagnac, P. (2014). Appearances can be deceptive: revealing a hidden viral infection with deep sequencing in a plant quarantine context. *PLoS ONE, 9,* e102945. DOI: 10.1371/journal.pone.0102945

Candresse, T., Lanneau, M., Revers, F., Grasseau, N., Macquaire, G., German, S., Malinovsky, T., & Dunez, J. (1995). An immunocapture PCR assay adapted to the detection and the analysis of the molecular variability of Apple chlorotic leaf spot virus. *Acta Horticulturae, 386,* 136-147. DOI : 10.17660/ActaHortic.1995.386.17

Candresse, T., Marais, A., Faure, C., & Gentit, P. (2013). Association of Little cherry virus 1 (LChV1) with the Shirofugen stunt disease and characterization of the genome of a divergent LChV1 isolate. *Phytopathology, 103*, 293-298. DOI: 10.1094/PHYTO-10-12-0275-R.

Candresse, T., Svanella-Dumas, L., & Le Gall, O. (2006). Characterization and partial genome sequence of stocky prune virus, a new member of the genus Cheravirus. *Archives of virology, 151*, 1179-1188. DOI: 10.1007/s00705-005-0682-y

Çelik, A., Santosa, A. I., Gibbs, A. J., & Ertunç, F. (2022). *Prunus necrotic ringspot virus* in Turkey: an immigrant population. *Archives of Virology, 167*, 553-562. DOI: 10.1007/s00705-022-05374-1.

Chen, S., Zhou, Y., Ye, T., Hao, L., Guo, L., Fan, Z., Li, S., & Zhou, T. (2014). Genetic variation analysis of apple chlorotic leaf spot virus coat protein reveals a new phylogenetic type and two recombinants in China. *Archives of Virology, 159*, 1431-1438. DOI: 10.1007/s00705-013-1927-9.

Clark, M. F., & Adams, A. N. (1977). Characteristics of the Microplate Method of Enzyme-Linked Immunosorbent Assay for the Detection of Plant Viruses. *Journal of General Virology, 34*, 475–483. DOI: 10.1099/0022-1317-34-3-475

Cochran, L. C., & Hutchins, L. M. (1941). A severe ring spot virosis in peach. *Phytopathology, 31*, 860.

Das, B., Ahmed, N., & Singh, P. (2011). *Prunus* diversity- early and present development: A review. *International Journal of Biodiversity and Conservation, 3*, 721-734. DOI: 10.5897/IJBC.9000118

De Andrade, R. R., & Vaslin, M. F. (2014). SearchSmallRNA: a graphical interface tool for the assemblage of viral genomes using small RNA libraries data. *Virology Journal 11*, 45. DOI: 10.1186/1743-422X-11-45.

De La Torre-Almaráz, R., Pallás, V., &Sánchez-Navarro, J. A. (2019). Molecular characterization of a new trichovirus from peach in Mexico. *Archives of Virology, 164*, 2617-2620. DOI : 10.1007/s00705-019-04358-y.

Desvignes, J. C. (1999). Virus Diseases of Fruit Trees. Paris, France, Centre Technique Interprofessionnel des Fruits et Legumes.

Desvignes, J. C., & Boye, R. (1988). Les différentes maladies provoquées par le Chlorotic leaf spot virus sur les arbres fruitiers. *Infos Ctifl, 45*, 31-34

Dias, N. P., Hu, R., Hensley, D. D., Hansen, Z., Domier, L., & Hajimorad, M. R. (2022). A Survey for Viruses and Viroids of Peach in Tennessee Orchards by RNA Sequencing. *Plant Health Progress*, First on line, DOI: 10.1094/PHP-01-22-0007-SC

Dodds, J. A., Morris, T. J., & Jordan, R. L. (1984). Plant viral double-stranded RNA. *Annual Review of Phytopathology, 22,* 151–168. DOI: 10.1146/annurev.py.22.090184.001055

Dunez, J., Delbos, R., Desvignes, J. C., Marenaud, C., Kuszala, J., & Vuittenez, A. (1970). Mise en evidence d'un virus de type Ring Spot sur Prunus cerasifera. Pages 117-128, *In*: VIIIth *European Symposium of Fruit Tree Virus Diseases Proceedings.*

Edgar, R. C, Taylor, J., Lin, V., Altman, T., Barbera, P., Meleshko, D., Lohr, D., Novakovsky, G., Buchfink, B., Al-Shayeb, B., Banfield, J. F., de la Peña, M., Korobeynikov, A., Chikhi, R., & Babaian, A. (2022). Petabase-scale sequence alignment catalyses viral discovery. *Nature 602,* 142-147. DOI: 10.1038/s41586-021-04332-2.

Elbeaino, T., Giampetruzzi, A., De Stradis, A., & Digiaro, M. (2014). Deep-sequencing analysis of an apricot tree with vein clearing symptoms reveals the presence of a novel betaflexivirus. *Virus research, 181,* 1-5. DOI: 10.1016/j.virusres.2013.12.030

Elena, S. F., Fraile, A., & García-Arenal, F. (2014). Evolution and Emergence of Plant Viruses. *Advances in Virus Research, 88,* 161–191. DOI: 10.1016/b978-0-12-800098-4.00003-9

François, S., Filloux, D., Fernandez, E., Ogliastro, M., & Roumagnac, P. (2018). Viral metagenomics approaches for high-resolution screening of multiplexed arthropod and plant viral communities. *Methods in Molecular Biology, 1746,* 77-95. DOI: 10.1007/978-1-4939-7683-6_7

Fox, A. (2020). Reconsidering causal association in plant virology. *Plant Pathology, 69,* 956–961 DOI: 10.1111/ppa.13199

Gaafar, Y. Z. A., Westenberg, M., Botermans, M., László, K., De Jonghe, K., Foucart, Y., Ferretti, L., Kutnjak, D., Pecman, A., Mehle, N., Kreuze, J., Muller, G., Vakirlis, N., Beris, D., Varveri, C., & Ziebell, H. (2021). Interlaboratory Comparison Study on Ribodepleted Total RNA High-Throughput Sequencing for Plant Virus Diagnostics and Bioinformatic Competence. *Pathogens, 10,* 1174. DOI: 10.3390/pathogens10091174.

Gaafar, Y. Z. A., & Ziebell, H. (2020). Comparative study on three viral enrichment approaches based on RNA extraction for plant virus/viroid detection using high-throughput sequencing. *PLoS ONE, 15,* e0237951. DOI: 10.1371/journal.pone.0237951

Gao, R., Xu, Y., Candresse, T., He, Z., Li, S., Ma, Y., & Lu, M. (2017). Further insight into genetic variation and haplotype diversity of Cherry virus A from China. *PloS ONE*, *12*, e0186273. DOI: 10.1371/journal.pone.0186273

García, J. A., Glasa, M., Cambra, M., & Candresse, T. (2014). Plum pox virus and sharka: a model potyvirus and a major disease. *Molecular Plant Pathology*, *15*, 226-241. DOI:10.1111/mpp.12083

Gentit, P., Delbos, R. P., Candresse, T., & Dunez, J. (2001). Characterization of a new nepovirus infecting apricot in Southeastern France: apricot latent ringspot virus. *European Journal of Plant Pathology, 107*, 485-494. DOI: 10.1023/A:1011274610537

Gildow, F., Damsteegt, V., Stone, A., Schneider, W., Luster, D., & Levy, L. (2004). Plum pox in North America: identification of aphid vectors and a potential role for fruit in virus spread. *Phytopathology*, *94*, 868-874. DOI: 10.1094/PHYTO.2004.94.8.868

Gilmer, R. M., Nyland, G., & Moore, J. D. (1976). Prune dwarf. Pages 104-132, In: *Virus diseases and noninfectious disorders of stone fruits in North America*. U.S. Agr. Hdbk.

Hadidi, A., & Barba, M. (2011). Economic impact of pome and stone fruit viruses and viroids. Pages 1-7, In: *Virus and virus-like diseases of pome and stone fruits*. *Virus and Virus-Like Diseases of Pome and Stone Fruits*, Hadidi A, Barba M, Candresse T, JelkmannW (eds.), APS Press, St Paul, Minnesota (USA).

Hamilton, A. J., & Baulcombe, D. C. (1999). A Species of Small Antisense RNA in Posttranscriptional Gene Silencing in Plants. *Science*, *286*, 950–952. DOI: 10.1126/science.286.5441.950

Harris, K. (1977). An ingestion-egestion hypothesis of noncirculative virus transmission by aphids. Pages 165-220, In: *Aphids as Virus Vectors.* K. Maramorosch (ed.) *Academic Press*, New York

He, Y., Cai, L., Zhou, L., Yang, Z., Hong, N., Wang, G., Li, Shifang, & Xu, W. (2017). Deep sequencing reveals the first fabavirus infecting peach. *Scientific Reports*, *7*, 1-11. DOI:10.1038/s41598-017-11743-7

Hily, J. M., Candresse, T., Garcia, S., Vigne, E., Tannière, M., Komar, V., Barnabé, G., Alliaume, A., Gilg, S., Hommay, G., Beuve, M., Marais, A. & Lemaire, O. (2018). High-Throughput Sequencing and the Viromic Study of Grapevine Leaves: From the Detection of Grapevine-Infecting Viruses to the Description of a New Environmental Tymovirales Member. *Frontiers in Microbiology*, *9*, 1782. DOI: 10.3389/fmicb.2018.01782

Hily, J. M., Poulicard, N., Candresse, T., Vigne, E., Beuve, M., Renault, L., Velt, A., Spilmonr, A. S., & Lemaire, O. (2020). Datamining, genetic diversity analyses, and

phylogeographic reconstructions redefine the worldwide evolutionary history of Grapevine Pinot gris virus and Grapevine berry inner necrosis virus. *Phytobiomes Journal*, *4*, 165-177. DOI: 10.1094/PBIOMES-10-19-0061-R

Ho, T., & Tzanetakis, I. E. (2014). Development of a virus detection and discovery pipeline using next generation sequencing. *Virology*, *471–473*, 54–60. DOI: 10.1016/j.virol.2014.09.019

Hou, Q., Han, T., Li, L., Wang, J., Yu, M., Zhang, S., Cao, M., & Yang, C. (2019). The complete nucleotide sequence and genome organization of a novel virus of the order Tymovirales isolated from *Prunus* davidiana (Carr.) Franch. in Liaoning, China. *Archives of Virology, 164,* 1245-1248. DOI: 10.1007/s00705-019-04220-1.

Idris, A., Al-Saleh, M., Piatek, M., Al-Shahwan, I., Ali, S., & Brown, J. (2014). Viral Metagenomics: Analysis of Begomoviruses by Illumina High-Throughput Sequencing. *Viruses*, *6*, 1219–1236. DOI: 10.3390/v6031219

Igori, D., Lim, S., Baek, D., Kim, S. Y., Seo, E., Cho, I. S., Choi, G. S, Lim, H. S, & Moon, J. S. (2017). Complete nucleotide sequence and genome organization of peach virus D, a putative new member of the genus Marafivirus. *Archives of Virology, 162*, 1769-1772. DOI: 10.1007/s00705-017-3255-y

Janick, J. (2005). The Origins of Fruits, Fruit Growing, and Fruit Breeding. *Plant Breeding Review 25,* 255-320.

James, D. & Jelkmann, W. (1998). Detection of cherry virus A in Canada and Germany. *Acta Horticulturae, 472,* 299-303. DOI: 10.17660/ActaHortic.1998.472.35

Jelkmann, W. (1995). Cherry virus A: cDNA cloning of dsRNA, nucleotide sequence analysis and serology reveal a new plant capillovirus in sweet cherry. *Journal of General Virology, 76*, 2015-2024. DOI: 10.1099/0022-1317-76-8-2015.

Jelkmann, W., & Kunze, L. (1995). Plum pseudopox in German prune after infection with an isolate of apple chlorotic leafspot virus causing plum line pattern. *Acta Horticulturae*, *386,* 122-125

Jeske, H. (2018). Barcoding of Plant Viruses with Circular Single-Stranded DNA Based on Rolling Circle Amplification. *Viruses*, *10*, 469. DOI: 10.3390/v10090469

Jo, Y., Yoo, S. H., Chu, H., Cho, J. K., Choi, H., Yoon, J. Y., Choi, S. K., & Cho, W.K. (2015). Complete Genome Sequences of Peach Latent Mosaic Viroid from a Single Peach Cultivar. *Genome Announcement*, *24*, e01098-15. DOI: 10.1128/genomeA.01098-15.

Jones, R. A. (2009). Plant virus emergence and evolution: Origins, new encounter scenarios, factors driving emergence, effects of changing world conditions, and prospects for control. *Virus Research*, *141,* 113–130. DOI: 10.1016/j.virusres.2008.07.028

Jones, R. A., & Naidu, R. A. (2019). Global dimensions of plant virus diseases: Current status and future perspectives. *Annual Review of Virology*, *6*, 387-409. DOI: 10.1146/annurev-virology-092818-015606

Katsiani, A.T., Maliogka, V. I., Candresse, T., Katis, N. I. (2014). Host-range studies, genetic diversity and evolutionary relationships of ACLSV isolates from ornamental, wild and cultivated Rosaceous species. *Plant Pathology*, *63,* 63-71. DOI : 10.1111/ppa.12058

Katsiani, A. T., Maliogka, V. I., Katis, N. I., Svanella-Dumas, L., Olmos, A., Ruiz-García, A. B., Marais, A., Faure, C., Theil, S., Lotos, L., & Candresse, T. (2018). High-Throughput Sequencing Reveals Further Diversity of Little Cherry Virus 1 with Implications for Diagnostics. *Viruses*, *21,* 385. DOI: 10.3390/v10070385.

Katsiani, A. T., Maliogka, V. I., Amoutzias, G. D., Efthimiou, K. E. & Katis, K. E. (2015). Insights into the genetic diversity and evolution of Little cherry virus 1. *Plant Pathology*, *64*, 817–824. DOI : 10.1111/ppa.12309

Keim-Konrad R & Jelkmann W. (1996). Genome analysis of the 3'-terminal part of the little cherry disease associated dsRNA reveals a monopartite clostero-like virus. *Archives of Virology, 141,* 1437-1451. DOI: 10.1007/BF01718246.

Kesanakurti, P., Belton, M., Saeed, H., Rast, H., Boyes, I., & Rott, M. (2016). Screening for plant viruses by next generation sequencing using a modified double strand RNA extraction protocol with an internal amplification control. *Journal of Virological Methods, 236,* 35–40. DOI: 10.1016/j.jviromet.2016.07.001

Khalili, M., Candresse, T., Faure, C., & Marais, A. (2020). Complete genome sequence of almond luteovirus 1, a novel luteovirus infecting almond. *Archives of Virology, 165,* 2123-2126. DOI: 10.1007/s00705-020-04715-2

Khalili, M., Candresse, T., Koloniuk, I., Safarova, D., Brans, Y., Faure, C., Delmas, M., Massart, S., Aranda, M. A., 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. (2022). The expanding menagerie of *Prunus*-infecting luteoviruses. *Phytopathology*. First look on line. DOI: 10.1094/PHYTO-06-22-0203-R.

Kinoti, W. M., Constable, F. E., Nancarrow, N., Plummer, K. M., Rodoni, B. (2017). Analysis of intra- host genetic diversity of *Prunus* necrotic ringspot virus (PNRSV) using amplicon next generation sequencing. *PLoS ONE, 12,* e0179284. DOI : 10.1371/journal.pone.0179284

Koloniuk, I., Sarkisova, T., Petrzik, K., Lenz, O., Přibylová, J., Fránová, J., Špak, J., Lotos, L., Beta, C., Katsiani, A., Candresse, T., & Maliogka, V. I. (2018). Variability studies of two *Prunus*-infecting fabaviruses with the aid of high-throughput sequencing. *Viruses, 10,* 204. DOI: 10.3390/v10040204

Kreuze, J. F., Perez, A., Untiveros, M., Quispe, D., Fuentes, S., Barker, I., & Simon, R. (2009). Complete viral genome sequence and discovery of novel viruses by deep sequencing of

small RNAs: A generic method for diagnosis, discovery and sequencing of viruses. *Virology*, *388*, 1–7. DOI: 10.1016/j.virol.2009.03.024

Kumar, A., Sharma, M. K., Wani, T. F., Sharma, A., & Nyorak, G. (2021). Varietal Wealth of *Prunus* Species, Pages 1-41, *In: Prunus - Recent Advances*, A. B. Küden, and A. Küden (Eds.). IntechOpen. DOI : 10.5772/intechopen.99048

Kutnjak, D., Rupar, M., Gutierrez-Aguirre, I., Curk, T., Kreuze, J. F., & Ravnikar, M. (2015). Deep sequencing of virus-derived small interfering RNAs and RNA from viral particles shows highly similar mutational landscapes of a plant virus population. *Journal of Virology*, *89*, 4760-4769. DOI: 10.1128/JVI.03685-14

Kutnjak, D., Tamisier, L., Adams, I., Boonham, N., Candresse, T., Chiumenti, M., De Jonghe, K., Kreuze, J. F., Lefebvre, M., Silva, G., Malapi-Wight, M., Margaria, P., Mavrič Pleško, I., McGreig, S., Miozzi, L., Remenant, B., Reynard, J. S., Rollin, J., Rott, M., Schumpp, O., Massart, S., & Haegeman, A. (2021). A Primer on the Analysis of High-Throughput Sequencing Data for Detection of Plant Viruses. *Microorganisms*, *9*, 841. DOI: 10.3390/microorganisms9040841.

Laurence, M., Hatzis, C., & Brash, D. E. (2014). Common contaminants in next-generation sequencing that hinder discovery of low-abundance microbes. *PloS ONE*, *9*, e97876. DOI: 10.1371/journal.pone.0097876

Lebas, B., Adams, I., Al Rwahnih, M., Baeyen, S., Bilodeau, G. J., Blouin, A. G., Candresse, T., Chandelier, A., De Jonghe, K., Fox, A., Gaafar, Y. Z. A., Gentit, P., Haegeman, A., Ho, W., Hurtado-Gonzales, O., Jonkers, W., Kreuze, J., Kutjnak, D., Landa, B., Liu, M., Maclot, F., Malapi-Wight, M., Maree, H. J., Martoni, F., Mehle, N., Minafra, A., Mollov, D., Moreira, A., Nakhla, M., Petter, F., Piper, A. M., Ponchart, J., Rae, R., Remenant, B., Rivera, Y., Rodoni, B., Roenhorst, J. W., Rollin, J., Saldarelli, P., Santala, J., Souza-Richards, R., Spadaro, D., Studholme, D. J., Sultmanis, S., van der Vlugt, R., Tamisier, L., Trontin, C., Vazquez-Iglesias, I., Vicente, C. S. L., Vossenberg, B. T. L. H., Wetzel, T., Ziebell, H., & Massart, S. (2022). Facilitating the adoption of high-

throughput sequencing technologies as a plant pest diagnostic test in laboratories: A step-by-step description. *EPPO Bulletin*, *52*, 394-418. DOI: 10.1111/epp.12863

Lebas, B. S. M., Elliott, D. R., VandenBrink, R., Ochoa-Corona, F. M., Tang, J. Z., & Alexander, B. J. R. (2004). Apple chlorotic leaf spot virus infection induces plum pox virus-like symptoms on plum in New Zealand. *Acta Horticulturae, 657,* 121-125.

Lefebvre, M., Theil, S., Ma Y., & Candresse T. (2019). The VirAnnot pipeline: a resource for automated viral diversity estimation and operational taxonomy units (OTU) assignation for virome sequencing data. *Phytobiomes Journal, 3*, 256-259. DOI: 10.1094/PBIOMES-07-19-0037-A

Lefeuvre, P., Martin, D. P., Elena, S. F., Shepherd, D. N., Roumagnac, P., & Varsani, A. (2019). Evolution and ecology of plant viruses. *Nature Reviews in Microbiology*, *17*, 632–644. DOI: 10.1038/s41579-019-0232-3

Leff, J. W., Lynch, R. C., Kane, N. C., & Fierer, N. (2017). Plant domestication and the assembly of bacterial and fungal communities associated with strains of the common sunflower, Helianthus annuus. *New Phytologist*, *214*, 412-423. DOI: 10.1111/nph.14323

Lenz, O., Přibylová, J., Fránová, J., Koloniuk, I., & Špak, J. (2017). Identification and characterization of a new member of the genus Luteovirus from cherry. *Archives of Virology, 162*, 587-590. DOI: 10.1007/s00705-016-3125-z

Liefting, L. W., Waite, D. W., & Thompson, J. R. (2021). Application of Oxford Nanopore Technology to Plant Virus Detection. *Viruses*, *13,* 1424. DOI: 10.3390/v13081424.

Liu, H., Fu, Y., Jiang, D., Li, G., Xie, J., Cheng, J., Peng, Y., Ghabrial, S. A., & Yi, X. (2010). Widespread horizontal gene transfer from double-stranded RNA viruses to eukaryotic nuclear genomes. *Journal of Virology*, *84*, 11876-11887. DOI: 10.1128/JVI.00955-10

Maclot, F., Candresse, T., Filloux, D., Malmstrom, C. M., Roumagnac, P., Van der Vlugt, R., & Massart, S. (2020). Illuminating an ecological blackbox: using high throughput

sequencing to characterize the plant virome across scales. *Frontiers in Microbiology*, *11*, 578064. DOI: 10.3389/fmicb.2020.578064

Maliogka, V., Minafra, A., Saldarelli, P., Ruiz-García, A., Glasa, M., Katis, N., & Olmos, A. (2018). Recent Advances on Detection and Characterization of Fruit Tree Viruses Using High-Throughput Sequencing Technologies. *Viruses*, *10*, 436. DOI : 10.3390/v10080436

Marais, A., Candresse, T., Svanella-Dumas, L., & Jelkmann, W. (2011). Cherry virus A, Pages 147-150, In: *Virus and Virus-Like Diseases of Pome and Stone Fruits*, Hadidi A, Barba M, Candresse T, JelkmannW (eds.), APS Press, St Paul, Minnesota (USA).

Marais, A., Faure, C., Bergey, B., & Candresse, T. (2018a). Viral Double-Stranded RNAs (dsRNAs) from Plants: Alternative Nucleic Acid Substrates for High-Throughput Sequencing. *Methods in Molecular Biology*, *1746*, 45–53. DOI: 10.1007/978-1-4939-7683-6_4

Marais, A., Faure, C., Couture, C., Bergey, B., Gentit, P., & Candresse, T. (2014). Characterization by deep sequencing of divergent Plum bark necrosis stem pitting-associated virus (PBNSPaV) isolates and development of a broad-spectrum PBNSPaV detection assay. *Phytopathology*, *104*, 660-666. DOI: 10.1094/PHYTO-08-13-0229-R

Marais, A., Faure, C., Mustafayev, E., & Candresse, T. (2015a). Characterization of new isolates of Apricot vein clearing-associated virus and of a new *Prunus*-infecting virus: Evidence for recombination as a driving force in Betaflexiviridae evolution. *PLoS ONE 10*, e0129469. DOI: 10.1371/journal.pone.0129469

Marais, A., Faure, C., Mustafayev, E., Barone, M., Alioto, D., & Candresse, T. (2015b). Characterization by deep sequencing of *Prunus* virus T, a novel Tepovirus infecting *Prunus* species. *Phytopathology*, *105*, 135-140. DOI: 10.1094/PHYTO-04-14-0125-R

Marais, A., Faure, C., Theil, S., & Candresse, T. (2018). Molecular characterization of a novel species of Capillovirus from Japanese apricot (*Prunus* mume). *Viruses*, *10*, 144. DOI: 10.3390/v10040144

Marais, A., Šafářová, D., Navrátil, M., Faure, C., Cornaggia, D., Brans, Y., ... & Candresse, T. (2020). Complete genome sequence of cherry virus T, a novel cherry-infecting tepovirus. *Archives of Virology, 165*, 1711-1714. DOI: 10.1007/s00705-020-04656-w

Marais, A., Svanella-Dumas, L., M. Barone, Gentit, P. Faure, C., Charlot, G., Ragozzino, A., Candresse T. (2012) Development of a polyvalent RT-PCR detection assay covering the genetic diversity of Cherry capillovirus A. *Plant Pathology*, *61*, 195-204. DOI: 10.1111/j.1365-3059.2011.02488.x

Maree, H. J., Fox, A., al Rwahnih, M., Boonham, N., & Candresse, T. (2018). Application of HTS for Routine Plant Virus Diagnostics: State of the Art and Challenges. *Frontiers in Plant Science*, *9*. DOI: 10.3389/fpls.2018.01082

Massart, S., Candresse, T., Gil, J., Lacomme, C., Predajna, L., Ravnikar, M., Reynard, J. S., Rumbou, A., Saldarelli, P., Škorić, D., Vainio, E. J., Valkonen, J. P. T., Vanderschuren, H., Varveri, C., & Wetzel, T. (2017). A Framework for the Evaluation of Biosecurity, Commercial, Regulatory, and Scientific Impacts of Plant Viruses and Viroids Identified by NGS Technologies. *Frontiers in Microbiology*, *8*. DOI: 10.3389/fmicb.2017.00045

Massart, S., Chiumenti, M., de Jonghe, K., Glover, R., Haegeman, A., Koloniuk, I., Komínek, P., Kreuze, J., Kutnjak, D., Lotos, L., Maclot, F., Maliogka, V., Maree, H. J., Olivier, T., Olmos, A., Pooggin, M. M., Reynard, J. S., Ruiz-García, A. B., Safarova, D., Candresse, T. (2019). Virus Detection by High-Throughput Sequencing of Small RNAs: Large-Scale Performance Testing of Sequence Analysis Strategies. *Phytopathology*, *109*, 488–497. DOI: 10.1094/phyto-02-18-0067-r

Massart, S., Lebas, B., Chabirand, A., Chappé, A. M., Dreo, T., Faggioli, F., Harrison, C., Macarthur, R., Mehle, N., Mezzalama, M., Petter, F., Ravnikar, M., Renvoisé, J. P., Spadaro, D., Tomassoli, L., Tomlinson, J., Trontin, C., van der Vlugt, R., Vučurović, A.,

Weekes, R., & Brostaux, Y. (2022). Guidelines for improving statistical analyses of validation datasets for plant pest diagnostic tests. *EPPO Bulletin, 52,* 419-433. DOI: 10.1111/epp.12862

Massart, S., Olmos, A., Jijakli, H., & Candresse, T. (2014). Current impact and future directions of high throughput sequencing in plant virus diagnostics. *Virus Research, 188*, 90–96. DOI: 10.1016/j.virusres.2014.03.029

Mathioudakis, M. M., Candresse, T., Katis, N. I. (2007). First Report of Apple chlorotic leaf spot virus in Quince in Greece. *Plant Disease*, *91*, 462. DOI: 10.1094/PDIS-91-4-0462A.

Matic, S., Minafra, A., Sánchez-Navarro, J. A., Pallás, V., Myrta, A., Martelli, G. P. (2009). 'Kwanzan Stunting' syndrome: detection and molecular characterization of an Italian isolate of Little cherry virus 1. *Virus Research*, *143*, 61-7. DOI: 10.1016/j.virusres.2009.03.005.

Mehetre, G. T., Leo, V. V., Singh, G., Sorokan, A., Maksimov, I., Yadav, M. K., Upadhyaya, K., Hashem, A., Alsaleh, A. N., Dawoud, T. M., Almaary, K. S., & Singh, B. P. (2021). Current Developments and Challenges in Plant Viral Diagnostics: A Systematic Review. *Viruses*, *13*, 412. DOI: 10.3390/v13030412

Milusheva, S., Phelan, J., Piperkova, N., Nikolova, V., Gozmanova, M., & James, D. (2019). Molecular analysis of the complete genome of an unusual virus detected in sweet cherry (*Prunus* avium) in Bulgaria. *European Journal of Plant Pathology*, *153*, 197-207. DOI: 10.1007/s00705-005-0503-3

Mink, G. I. and Shay, J. R. (1959). Preliminary evaluation of some Russian apple varieties as indi-cators for apple viruses. Suppl. *P1. Disease Report, 245,* 3-17.

Molnár, A., Csorba, T., Lakatos, L., VáRallyay, E., Lacomme, C., & Burgyán, J. (2005). Plant Virus-Derived Small Interfering RNAs Originate Predominantly from Highly Structured Single-Stranded Viral RNAs. *Journal of Virology, 79,* 7812–7818. DOI : 10.1128/jvi.79.12.7812-7818.2005

Moreno, A. B. & Lopez-Moya J. J. (2020). When viruses play team sports: Mixed infection in plants. *Phytopathology, 110,* 29-48. DOI : 10.1094/PHYTO-07-19-0250-FI

Moubset, O., François, S., Maclot, F., Palanga, E., Julian, C., Claude, L., ... & Roumagnac, P. (2022). Virion-associated nucleic acid-based metagenomics: a decade of advances in molecular characterization of plant viruses. *Phytopathology*, On line First. DOI: 10.1094/PHYTO-03-22-0096-RVW .

Myrta, A., Matic, S., Malinowski, T., Pasquini G., & Candresse T. (2011). Apple chlorotic leaf spot virus in Stone fruits. Pages 85-90, In: *Virus and virus-like diseases of pome and stone fruits*, Hadidi A, Barba M, Candresse T, JelkmannW (eds.), APS Press, St Paul, Minnesota (USA).

Nakahara, K. S., Yoshida, K., Suzaki, K., Yoshikawa, N., Ito, T. (2011). Sensitive PCR based detection of Apple chlorotic leaf spot virus heterogeneous in apple trees. *Japan Agricultural Research Quarterly, 45,* 411–421. DOI: 10.6090/jarrq.45.411

Nemeth, M. A. R. I. A. (1986). Virus, mycoplasma and Rickettsia diseases of fruit trees. Pages 463-479, In: *Virus, Mycoplasma and Rickettsia Diseases of Fruit Trees*, Martinus Nijhoff Publishers Dordrech.

Ng, T. F. F., Willner, D. L., Lim, Y. W., Schmieder, R., Chau, B., Nilsson, C., Anthony, S., Ruan, Y., Rohwer, F., & Breitbart, M. (2011). Broad Surveys of DNA Viral Diversity Obtained through Viral Metagenomics of Mosquitoes. *PLoS ONE, 6,* e20579. DOI: 10.1371/journal.pone.0020579

Nourinejhad Zarghani, S., Hily, J. M., Glasa, M., Marais, A., Wetzel, T., Faure, C., Vigne, E., Velt, A., Lemaire, O., Boursiquot, J. M., Okic, A., Ruiz-Garcia, A. B., Olmos, A., Lacombe, T., & Candresse, T. (2018). Grapevine virus T diversity as revealed by full-length genome sequences assembled from high- throughput sequence data. *PLoS ONE, 130,* e0206010. DOI: /10.1371/journal. pone.0206010

Oberemok, V., Laikova, K., Golovkin, I., Kryukov, L., & Kamenetsky-Goldstein, R. (2021). Biotechnology of virus eradication and plant vaccination in phytobiome context. *Plant Biology*, *24*, 3–8. DOI: 10.1111/plb.13338

Ohan, N. W., & Heikkila, J. J. (1993). Reverse transcription-polymerase chain reaction: An overview of the technique and its applications. *Biotechnology Advances*, *11*, 13–29. DOI: 10.1016/0734-9750(93)90408-f

Orfanidou, C. G., Xing, F., Zhou, J., Li, S., Katis, N., & Maliogka, V. I. (2021). Identification and Sequence Analysis of a Novel Ilarvirus Infecting Sweet Cherry. *Plants*, *10*, 514. DOI : 10.3390/plants10030514

Pallas, V., Aparicio, F., Herranz, M. C., Amari, K., Sanchez-Pina, M. A., Myrta, A., & Sanchez-Navarro, J. A. (2012). Ilarviruses of *Prunus* spp.: a continued concern for fruit trees. *Phytopathology*, *102*, 1108-1120. DOI: 10.1094/PHYTO-02-12-0023-RVW

Pecman, A., Adams, I., Gutiérrez-Aguirre, I., Fox, A., Boonham, N., Ravnikar, M., & Kutnjak, D. (2022). Systematic Comparison of Nanopore and Illumina Sequencing for the Detection of Plant Viruses and Viroids Using Total RNA Sequencing Approach. *Frontiers in Microbiology*, *13*, 883921. DOI: 10.3389/fmicb.2022.883921

Pecman, A., Kutnjak, D., Gutiérrez-Aguirre, I., Adams, I., Fox, A., Boonham, N., & Ravnikar, M. (2017). Next Generation Sequencing for Detection and Discovery of Plant Viruses and Viroids: Comparison of Two Approaches. *Frontiers in Microbiology*, *8*. DOI: 10.3389/fmicb.2017.01998

Pooggin, M. M. (2018). Small RNA-Omics for Plant Virus Identification, Virome Reconstruction, and Antiviral Defense Characterization. *Frontiers in Microbiology*, *9*. DOI: 10.3389/fmicb.2018.02779

Potter, D., Eriksson, T., Evans, R. C., Oh, S., Smedmark, J. E. E., Morgan, D. R., Kerr, M., Robertson, K. R., Arsenault, M., Dickinson, T. A., & Campbell, C. S. (2007). Phylogeny

and classification of Rosaceae. *Plant Systematics and Evolution*, *266*, 5–43. DOI: 10.1007/s00606-007-0539-9

Rehder, A. (1940). Manual of cultivated trees and shrubs hardy in North America, exclusive of the subtropical and warmer temperate regions, (Enlarged 2nd ed.). The Macmillan Company.

Ren, J., Ahlgren, N.A., Lu, Y.Y., Fuhrman, J.A., & Sun, F. (2017). VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. *Microbiome, 5*, 1-20. DOI: 10.1186/s40168-017-0283-5

Reyes, A., Semenkovich, N. P., Whiteson, K., Rohwer, F., & Gordon, J. I. (2012). Going viral: next-generation sequencing applied to phage populations in the human gut. *Nature Reviews in Microbiology 10,* 607–617. DOI: 10.1038/nrmicro2853

Rodoni, B. (2009). The role of plant biosecurity in preventing and controlling emerging plant virus disease epidemics. *Virus research*, *141*, 150-157. DOI: 10.1016/j.virusres.2008.11.019

Roossinck MJ. (2019). Viruses in the phytobiome. *Current Opinion in Virology*, *37*, 72-76. DOI : 10.1016/j.coviro.2019.06.008.

Roossinck, M. J., Martin, D. P., & Roumagnac, P. (2015). Plant virus metagenomics: advances in virus discovery. *Phytopathology*, *105*, 716-727. DOI: 10.1094/PHYTO-12-14-0356-RVW

Roossinck, M. J., Saha, P., Wiley, G. B., Quan, J., White, J. D., Lai, H., Chavarría, F., Shen, G., & Roe, B. A. (2010). Ecogenomics: using massively parallel pyrosequencing to understand virus ecology. *Molecular Ecology*, *19*, 81-88. DOI: 10.1111/j.1365-294X.2009.04470.x

Roumagnac, P., Mollov, D., Daugrois, J. H., & Filloux, D. (2018). Viral metagenomics and sugarcane pathogens. Pages 183-200, In: *Achieving Sustainable Cultivation of*

Sugarcane. Volume 2: Breeding, pests and diseases, Rott Philippe (ed.). Cambridge: Burleigh Dodds Science Publishing. DOI: 10.19103/as.2017.0035.20

Rubio, L., Galipienso, L., & Ferriol, I. (2020). Detection of Plant Viruses and Disease Management: Relevance of Genetic Diversity and Evolution. *Frontiers in Plant Science*, *11*. DOI: 10.3389/fpls.2020.01092

Rwahnih, M. A., Dave, A., Anderson, M. M., Rowhani, A., Uyemoto, J. K., & Sudarshana, M. R. (2013). Association of a DNA virus with grapevines affected by red blotch disease in California. *Phytopathology*, *103*, 1069-1076. DOI: 10.1094/PHYTO-10-12-0253-R

Sabanadzovic, S., Ghanem-Sabanadzovic, N. A., Rowhani, A., Grant, J. A., & Uyemoto, J. K. (2005). Detection of cherry virus A, cherry necrotic rusty mottle virus and little cherry virus 1 in California orchards. *Journal of Plant Pathology, 87*, 173–177.

Šafářová, D., Faure, C., Candresse, T., Navrátil, M., Nečas, T., & Marais, A. (2017). First report of Little cherry virus 1 infecting apricot in the Czech Republic. *Plant Disease*, *101*, 845. DOI: 10.1094/PDIS-09-16-1289-PDN

Salem, N., Mansour, A., Al-Musa, A., Al-Nsour, A., & Hammond, R. (2004). Identification and partial characterization of Prunus necrotic ringspot virus on stone fruits in Jordan. *Journal of Plant Pathology*, *86*, 85-90. DOI:10.4454/JPP.V86I1.941

Schmieder, R., & Edwards, R. (2011). Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PloS ONE, 6*, e17288. DOI: 10.1371/journal.pone.0017288

Scholthof, K. B. G., Adkins, S., Czosnek, H., Palukaitis, P., Jacquot, E., Hohn, T., ... & Foster, G. D. (2011). Top 10 plant viruses in molecular plant pathology. *Molecular Plant Pathology, 12*, 938-954. DOI: 10.1111/j.1364-3703.2011.00752.x

Seguin J, Otten P, Baerlocher L, Farinelli L, Pooggin MM. (2014a) MISIS: a bioinformatics tool to view and analyze maps of small RNAs derived from viruses and genomic loci

generating multiple small RNAs. *Journal of Virological Methods, 195*, 120-122. DOI: 10.1016/j.jviromet.2013.10.013.

Seguin, J., Rajeswaran, R., Malpica-Lopez, N., Martin, R. R., Kasschau, K., Dolja, V. V., ... & Pooggin, M. M. (2014b). De novo reconstruction of consensus master genomes of plant RNA and DNA viruses from siRNAs. *PloS ONE*, *9*, e88513. DOI: 10.1371/journal.pone.0088513

Sidharthan, V. K., Kalaivanan, N. S., & Baranwal, V. K. (2021). Discovery of putative novel viruses in the transcriptomes of endangered plant species native to India and China. *Gene, 786*, 145626. DOI: 10.1016/j.gene.2021.145626

Sokhandan-Bashir N, Kashiha M, Koolivand D, Eini O. (2017). Detection and phylogenetic analysis of *Prunus* necrotic ringspot virus isolates from stone fruits in Iran. *Journal of Plant Pathology, 99*, 1–14. DOI: 10. 4454/ jpp. v99i3. 3986

Soltani, N., Stevens, K. A., Klaassen, V., Hwang, M. S., Golino, D. A., & Al Rwahnih, M. (2021). Quality assessment and validation of high-throughput sequencing for grapevine virus diagnostics. *Viruses*, *13*, 1130. DOI: 10.3390/v13061130

Spiegel, S., Rosner, A., Tam, Y., Zilkah, S., Faingersh, E., Rotbaum, A. and Krizbai, L. (1998). Detection of Prune dwarf virus in sweet cherry in Israel. *Acta Horticulturae, 472*, 249-256. DOI: 10.17660/ActaHortic.1998.472.27

Svanella-Dumas, L., Marais, A., Gentit, P., Lamorte, J., & Candresse, T. (2005). First report on the natural occurrence of Cherry virus A in Mirabelle plum (Prunus domestica var. insititia). *Plant Disease*, *89*, 433-433. DOI: 10.1094/PD-89-0433A.

Sukhorukov, G., Khalili, M., Gascuel, O., Candresse, T., Marais, A., Nikolski, M. (2022). VirHunter: a deep learning-based method for detection of novel RNA viruses in plant sequencing data. *Frontiers in Bioinformatics*, *2*, 867111. DOI: 10.3389/fbinf.2022.867111

Tampuu, A., Bzhalava, Z., Dillner, J., & Vicente, R. (2019). ViraMiner: Deep learning on raw DNA sequences for identifying viral genomes in human samples. *PLoS ONE, 14*, e0222271. DOI: 10.1371/journal.pone.0222271

Uyemoto, J. K., & Scott, S. W. (1992). Important diseases of *Prunus* caused by viruses and other graft-transmissible pathogens in California and South Carolina. *Plant Disease, 76*, 5-11. DOI: 10.1094/PD-76-0005

Villamor, D. E. V., Ho, T., Al Rwahnih, M., Martin, R. R., & Tzanetakis, I. E. (2019). High throughput sequencing for plant virus detection and discovery. *Phytopathology*, *109*, 716-725. DOI: 10.1094/PHYTO-07-18-0257-RVW.

Villamor, D. E. V., Mekuria, T. A., Pillai, S. S., & Eastwell, K. C. (2016). High-throughput sequencing identifies novel viruses in nectarine: insights to the etiology of stem-pitting disease. *Phytopathology*, *106*, 519-527. DOI: 10.1094/PHYTO-07-15-0168-R

Villamor, D. E. V., Pillai, S. S., & Eastwell, K. C. (2017). High throughput sequencing reveals a novel fabavirus infecting sweet cherry. *Archives of Virology*, *162*, 811-816. DOI: 10.1007/s00705-016-3141-z

Visser, M., Bester, R., Burger, J. T., and Maree, H. J. (2016). Next-generation sequencing for virus detection: covering all the bases. *Virology Journal, 13*, 85. DOI: 10. 1186/s12985-016-0539-x

Vivek, A., Zahra, S., & Kumar, S. (2020). From current knowledge to best practice: A primer on viral diagnostics using deep sequencing of virus-derived small interfering RNAs (vsiRNAs) in infected plants. *Methods*, *183*, 30–37. DOI: 10.1016/j.ymeth.2019.10.009

Wommack, K. E., Bhavsar, J., Polson, S. W., Chen, J., Dumas, M., Srinivasiah, S., Furman, M., Jamindar, S., Nasko, D. J. (2012) VIROME: a standard operating procedure for analysis of viral metagenome sequences. *Stand Genomic Sciences 6*, 427-39. DOI: 10.4056/sigs.2945050

Wren, J. D., Roossinck, M. J., Nelson, R. S., Scheets, K., Palmer, M. W., & Melcher, U. (2006). Plant virus biodiversity and ecology. *PLoS Biology, 4*, e80. DOI: 10.1371/journal.pbio.0040080.

Wu, L., Liu, H., Bateman, M., Komorowska, B., & Li, R. (2019). First identification and molecular characterization of a novel cherry robigovirus. *Archives of Virology*, *64*, 3103-3106. DOI: 10.1007/s00705-019-04401-y

Wu, L. P., Liu, H. W., Bateman, M., Liu, Z., & Li, R. (2017). Molecular characterization of a novel luteovirus from peach identified by high-throughput sequencing. *Archives of Virology*, *162*, 2903-2905. DOI: 10.1007/s00705-017-3388-z

Wu, Q., Ding, S. W., Zhang, Y., & Zhu, S. (2015). Identification of Viruses and Viroids by Next-Generation Sequencing and Homology-Dependent and Homology-Independent Algorithms. *Annual Review of Phytopathology*, *53*, 425–444. DOI: 10.1146/annurev-phyto-080614-120030

Yaegashi, H., Yoshikawa, N., & Candresse, T. (2011). Apple chlorotic leaf spot virus in pome fruits. Pages 17-22, In *Virus and Virus-Like Diseases of Pome and Stone Fruits*, Hadidi A, Barba M, Candresse T, JelkmannW (eds.), APS Press, St Paul, Minnesota (USA).

Zhang, C., Wu, Z., Li, Y., & Wu, J. (2015). Biogenesis, function, and applications of virus-derived small RNAs in plants. *Frontiers in Microbiology*, *6*, 1237. DOI: 10.3389/fmicb.2015.01237

Zhang, Z., Qi, S., Tang, N., Zhang, X., Chen, S., Zhu, P., Ma, L., Cheng, J., Xu, Y., Lu, M., Wang, H., Ding, S. W., Li, S., &Wu, Q. (2014). Discovery of replicating circular RNAs by RNA-seq and computational algorithms. *PLoS Pathogen, 10*, e1004553. DOI: 10.1371/journal.ppat.1004553.

Zheng, Y., Gao, S., Padmanabhan, C., Li, R., Galvez, M., Gutierrez, D., Fuentes, S., Ling, K. S., Kreuze, J., & Fei, Z. (2017). VirusDetect: An automated pipeline for efficient virus

discovery using deep sequencing of small RNAs. *Virology, 500,* 130-138. DOI: 10.1016/j.virol.2016.10.017.

Zhou, J., Cao, K., Zhang, Z., Wang, L., & Li, S. Identification and characterization of a novel rhabdovirus infecting peach in China. (2020). *Virus Research, 280,* 197905. DOI: 10.1016/j.virusres.2020.197905.

Zhou, J., Zhang, Z., Lu, M., Xiao, H., Habili, N., & Li, S. (2018). Complete nucleotide sequence of a new virus, peach chlorotic leaf spot virus, isolated from flat peach in China. *Archives of Virology*, *163*, 3459-3461. DOI: 10.1007/s00705-018-3984-6