HAL open science

Analysis of the virome and reciprocal transfers of viruses between cultivated and wild carrot populations

Deborah Schönegger

To cite this version:

Deborah Schönegger. Analysis of the virome and reciprocal transfers of viruses between cultivated and wild carrot populations. Agricultural sciences. Université de Bordeaux, 2023. English. NNT : 2023BORD0021 . tel-04033935

HAL Id: tel-04033935

https://theses.hal.science/tel-04033935v1

Submitted on 17 Mar 2023

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 doctoral Science de la Vie et de la Santé

Mention Biologie Végétale

Par

Deborah SCHÖNEGGER

Analyse du virome et des flux de virus entre populations de carottes cultivées et sauvages

Sous la direction de : **Thierry CANDRESSE**

Soutenue le 30 Janvier 2023

Membres du jury :

M. Thierry CANDRESSE, Directeur de recherche, INRAE (VILLENAVE D'ORNON), ***Directeur de thése***
M. Neil BOONHAM, Professeur de University of Newcastle (Newcastle Upon Tyne, ROYAUME-UNI), Rapporteur
Mme Veronique BRAULT, Directrice de recherche, INRAE (Colmar), Rapporteure
Mme Anna-Liisa LAINE, Professeur, University of Zurich (Zurich, SUISSE), Examinatrice
Mme Marie-Helene OGLIASTRO, Directrice de recherche, INRAE (Montpellier), Presidente du jury

Analyse du virome et des flux de virus entre populations de carottes cultivées et sauvages

Résumé

Les technologies de séquençage haut débit (HTS pour High-Throughput Sequencing) appliquées aux études de métagénomiques ont révolutionné notre vision de la biodiversité virale, avec un impact notable dans les domaines du phytodiagnostic et de l'écologie virale, permettant d'identifier tous les virus présents dans un échantillon sans connaissance préalable. Cependant, comme pour toute technologie, les stratégies HTS et les pipelines d'analyses des données générées doivent être validés pour garantir leur fiabilité. Ces évaluations sont basées sur la comparaison des performances théoriques et réelles, en utilisant des communautés virales de composition connue. Néanmoins, de telles études font encore largement défaut dans le domaine de la virologie végétale. Au cours de cette thèse, les défis méthodologiques et les questions écologiques liées à la description et à l'analyse des communautés de virus végétaux ont été abordés. Tout d'abord, des communautés synthétiques de phytovirus de complexités variables ont été utilisées pour évaluer les deux stratégies d'enrichissement des virus les plus utilisées dans les études de métagénomique, à savoir l'analyse des acides nucléiques associés aux virions (VANA, Virion-Associated Nucleic Acids) et l'analyse des ARN double brin (ARNdb). Nos résultats montrent que l'approche de l'ARNdb fournit systématiquement une description plus complète du virome à ARN mais, comme attendu, est peu performante pour les virus à ADN. Par ailleurs, on a pu mettre en évidence une corrélation robuste entre la profondeur de séquençage des échantillons et l'exhaustivité de la description du virome, ainsi que l'influence d'autres paramètres tels que la complexité du virome, l'utilisation de contigs et la longueur minimale des contigs. Dans une approche plus orientée vers des questions d'écologie, le virome de 45 populations de carottes sauvages (*Daucus carota ssp. carota*) et cultivées (*Daucus carota* ssp. *sativus*) récoltées en France et de six populations collectées en Espagne a été analysé en utilisant une approche HTS basée sur l'ARNdb. Les résultats ont permis de constater que: (i) les carottes présentent un virome particulièrement diversifié, riche en nouveaux virus qui ont ensuite été caractérisés au niveau moléculaire et (ii) le virome des carottes cultivées est très différent de celui des carottes sauvages. Les signatures virales entre ces deux populations montrent une présence différentielle de certains virus, mais aussi une prévalence différentielle d'autres virus. Sur la base de l'analyse de plantes individuelles, la diversité génétique des virus transmis par pucerons a été analysée, mettant en évidence une différenciation intraspécifique des populations virales entre hôtes sauvages et cultivés pour plusieurs de ces virus.

Il s'agit d'un résultat particulièrement intéressant si l'on considère que les carottes cultivées et sauvages sont membres de la même espèce, et qu'elles représentent donc un pathosystème pour lequel peu de contraintes génétiques et botaniques sont attendues sur le flux de virus entre ces deux populations. Globalement, les résultats obtenus suggèrent un impact majeur des conditions de culture liées à l'agriculture dans la structuration de la richesse et de la composition du virome. Ils ont également montré l'existence de flux viraux entre les compartiments sauvage et cultivé mais aussi l'existence de barrières limitant ou empêchant les échanges de certains virus ou de certains groupes d'isolats viraux. Le pathosystème carotte utilisé ici a permis de mieux comprendre les viromes contrastés qui peuvent s'assembler chez des plantes étroitement apparentées à l'interface agro-écologique et fournissent une base pour des recherches plus approfondies sur les barrières biologiques et écologiques s'appliquant aux mouvements des virus, ainsi que sur les interactions entre virus transmis par pucerons dans un réseau complexe d'interdépendances mutuelles.

Mots clés: virome, carotte, séquençage haut débit, flux de virus, diversitè

Analysis of the virome and reciprocal transfers of viruses between wild and cultivated carrot populations

Abstract

HTS-based metagenomics has revolutionized our view of biodiversity and has become a powerful tool in plant virus diagnostics and ecology, allowing to theoretically capture without prior knowledge all viruses present in a sample. However, as with any technology, HTS strategies and the corresponding data processing pipelines need to be validated to provide reliable assessments of plant health status and balance the cost of sequencing with the specificity and sensitivity levels required by the study. In this context, metagenomics benchmarking efforts are often based on mock communities of known composition thus allowing to compare the actual *vs* the expected performance. Despite their importance, such comparative studies in virology are mainly limited to the clinical sector and are largely lacking in the plant virology field. In the present thesis both methodological challenges and ecological questions in relation to the description and analysis of plant virus communities have been addressed. Phytovirus synthetic communities of varying complexity were used to benchmark the two most frequently used virus enrichment strategies for plant virus metagenomics, virion associated nucleic acids (VANA) and double stranded (ds)RNA. The results show that the dsRNA approach systemically provided a more complete description of the RNA virome but, as expected, performed poorly for DNA viruses. A robust correlation between sample sequencing depth and the completeness of virome description was observed, together with the influence of other parameters such as virome complexity, use of *de novo* assembled contigs and minimal contig length. In a second more ecologically oriented approach the virome of 45 different wild (*Daucus carota* ssp. *carota*) and cultivated carrot (*Daucus carota* ssp. *sativus*) populations in France and six additional populations in Spain was analyzed using a dsRNA-based HTS approach. The datasets generated have revealed two important findings: (i) carrots exhibit a particularly diverse virome, rich in novel viruses that were then molecularly characterized and (ii) virus communities differ remarkably between wild and cultivated carrots. The contrasting virome fingerprints between field and wild carrot populations involve both differential presence of some viruses and differential prevalence of other viruses. Based on analysis of individual plants by HTS sequencing and RT-PCR, population genetics of aphid-transmitted viruses were analyzed, highlighting intraspecific viral population differentiation between wild and cultivated hosts for several of them. This is a particularly interesting result considering that both cultivated and wild carrots are members of the same species, and thus

represent a pathosystem with expected least genetic and botanical constraints on potential virus flow between a crop and its weedy relative. Taken together, the results obtained suggest a major impact of growth conditions created by agriculture in shaping virome richness and composition. They demonstrated the existence of viral fluxes between the wild and cultivated compartments but also the existence of barriers limiting or altogether preventing exchange of some viruses or of some groups of isolates. The carrot pathosystem used here allowed to gain deep insights into the contrasted viromes that can assemble in closely related plants a the agroecological interface and to provide a baseline for further investigations of the biological and ecological barriers to virus movements and on the complex interaction linking aphid-transmitted viruses in a complex network of mutual interdependencies.

Keywords: virome, carrot, high throughput sequencing, virus fluxes, diversity

---

Unité de recherche

UMR 1332 Biologie du Fruit et Pathologie, INRA & Université de Bordeaux, Campus INRA de la Grande Ferrade, 71 avenue Edouard Bourleaux, CS20032, 33882 VILLENAVE d'ORNON cedex

Acknoledgments

First and foremost, I would like to thank my supervisor, **Dr. Thierry Candresse**. He patiently guided me through this PhD and brought many innovative ideas to our discussions. His 100% commitment to the project and his continuous help whenever I had questions or doubts contributed greatly to the success of this work. His immense knowledge, great enthusiasm for science, and collaborative spirit remain a great inspiration to me.

I would like to express my sincere gratitude to my co-supervisor **Dr. Armelle Marais**. She helped me a lot during this PhD, whether it was organizing the lab work to be 100% efficient or clarifying doubts whenever I felt unsure. She carefully revised my manuscripts and always had an open ear when I was worried about work or my personal life.

I would like to thank **Dr. Fernando Garcia-Arenal, Dr. Phillipe Roumagnac, and Dr. Emmanuel Geoffriau**, who were part of the committee for the midterm examination of my dissertation and gave me valuable suggestions in our discussions to broaden my horizons and improve my work. Thankfully, I was also able to spend a month in Madrid at CBGP in the lab of Fernando Garcia Arenal, which was very important as I learned many new data analysis strategies. I was able to perform important experiments in the lab of Dr. Phillipe Roumagnac, who welcomed me warmly in Montpellier, and finally I received many additional samples from Emmanuel Roumagnac, which allowed me to perform interesting experiments.

I would like to express my deepest gratitude to the members of my dissertation committee. I thank **Prof. Neil Boonham** and **Dr. Veronique Brault** for evaluating my dissertation manuscript and, of course, **Prof. Anna-Liisa Laine** and **Dr. Marie-Helene Ogliastro** for participating in the jury, for the interesting discussion and positive feedback during my dissertation defense that made this day an unforgettable success.

My huge thanks go to **Chantal Faure**, who was always friendly and helpful in the lab and during sampling. I thank **Laurence Svanella-Dumas** and **Sylvie Germana** for their emotional support whenever I was homesick. I would like to thank Marie Lefebvre who helped me analyze the data and was so kind and helpful.

My sincere thanks also go to all my dear colleagues in the virology team for their open-mindedness and all the humorous conversations in the coffee corner that made the daily life more pleasant. I would like to take this opportunity to thank office and PhD colleagues **Maryam, Mingshuo,**

Justine, Julien, Amélie, Alexandre, Loïse and Fares who formed a mixed and international group with whom I spent many funny moments.

I am thankful for the funding of this project by European Union through Horizon 2020 Marie Sklodowska-Curie Actions Innovative Training Network (H2020 MSCA-60 INT) project "INEXTVIR" (Grant agreement number 813542)

Last but not least, I would like to thank all **INEXTVIR members** for organizing many networking meetings that allowed an informative scientific but also personal exchange of experiences. We have spent many beautiful and unique moments together, especially after the pandemic when we could meet in person. I especially thank our dear advisor **Calolyn Malmstrom** for organizing a journal club with stimulating and enthusiastic discussions on newest topics in science.

Finally, I deeply thank my parents who support me in every choice I take in my life and always encourage me. I endlessly thank my boyfriend Lorenzo who is always on my side and supports me in every aspect of my life.

SCIENTIFIC PRODUCTION

List of publications

- **Schönegger D**, Moubset O, Margaria P, Menzel W, Winter S, Roumagnac P, Marais A, Candress T (2022) Benchmarking of virome metagenomic analysis approaches using a large scale, 60+ members, viral synthetic community. ***To be submitted***
- **Schönegger D**, Marais A, Babalola B, Faure C, Lefebvre L, Svanella-Dumas L, Garcia Arenal F, Brázdová S, Candresse T (2022) Carrot populations in France and Spain host a complex virome rich in previously uncharacterized viruses. ***To be submitted***
- **Schönegger D**, Marais A, Garcia-Arenal F, Martinez Jimenez M, McLeish M, Lefebvre M, Faure C, Svanella-Dumas L, Candresse T (2022) Metagenomic analysis of cultivated and wild carrots reveals different virome fingerprints and provides insights into viral fluxes between carrot populations. ***To be submitted***
- **Schönegger D**, Marais A, Faure C, Candresse T (2022) A new flavi-like virus identified in populations of wild carrots. Arch. Virol. 167, 2407-2409. https://doi: 10.1007/s00705-022-05544-1
- **Schönegger D**, Babalola BM, Marais A, Faure C, Candresse T (2022) Diversity of polerovirus-associated RNAs in the virome of wild and cultivated carrots. Plant. Pathol. 71, 1892– 1900. https://doi.org/10.1111/ppa.13623
- Babalola BM, **Schönegger D**, Faure C, Marais A, Fraile A, Garcia-Arenal F, Candresse T (2022) Identification of two novel putative satellite RNAs with hammerhead structures in the virome of French and Spanish carrot samples. Arch. Virol. 167, 2287-2292. https://doi.org/10.1007/s00705-022-05538-z

List of Oral Communications

- MAY 2021, Journée Scientifique de l'Ècole Doctoral SVS, Bordeaux, France
- SEP 2021, 18éme Rencontres de Virologie Végetale, Aussois, France
  Jury prize 'coup de coeur'
- JUN 2022, 12éme Journée des Doctorants SPE, Bordeaux, France
- JUN 2022, 3rd Microbiology Day, Bordeaux, France
- OCT, 2022 International Advances in Plant Virology 2022, Ljubljana, Slovenia

Formations suives

Catégorie 1 : Connaître l'environnement académique

- Research integrity in scientific professions

  15 heures enregistrées par: Sciences de la Vie et de la Santé

Catégorie 2 : Mener à bien ses activités de recherche

- Èthique de la recherche (18 mai 2021 – 15 juin 2021) Université de Lyon

  12 heures enregistrées par: Sciences de la Vie et de la Santé
- Training school 1 - Marie Sklodowska-Curie Innovative Training Network (MSCA-ITN) (21 avril 2020) National institute of biology - Ljubljana (Slovenia)

  12 heures enregistrées par : Sciences de la Vie et de la Santé.

Catégorie 3 : Communiquer en langue étrangère

- Français Langue Etrangère (Autumn 2020)

  38 heures
- Français Langue Etrangère (Spring 2020) à distance plateforme ZOOM

  40 heures

Total participation : 117 heures/5 modules

Contents

GENERAL INTRODUCTION 1

1 Plant viruses – definition, classification and diversity 1

2 Plant virology - Historical perspective and methodological advances 2

2.1 HTS based metagenomics in plant virus diagnostics and ecology 4

2.2 The plant virome and its technical challenges 5

3 Plant virus ecology and evolution 7

3.1 The ecological trinity 8

3.1.1 Virus-vector interactions 8

3.1.2 Virus-host interactions 9

3.1.3 Virus-host-vector interactions 10

3.2 Viruses at the agroecological interface 11

4 Carrots and their viruses 12

4.1 Biology and life cycles of *Daucus carota* L. 12

4.2 Carrot viruses 14

4.3 Carrot as model plant for virome comparisons 18

5 Outline of the scientific tasks and questions addressed in the present thesis 18

CHAPTER I 20

Manuscript: *"Benchmarking of virome metagenomic analysis approaches using a large, 60+ members, viral synthetic community"* 21

Abstract 22

INTRODUCTION 23

MATERIALS AND METHODS 26

RESULTS 30

DISCUSSION 38

REFERENCES 43

Tables and Figures 50

Supplementary Materials [tables are available at https://doi.org/10.57745/XYIKFK] 61

CHAPTER II **68**

Manuscript: "*Carrot populations in France and Spain host a complex virome rich in previously uncharacterized viruses"* **69**

Abstract 70

INRTODUCTION 71

MATERIALS AND METHODS 73

RESULTS 76

DISCUSSION 90

REFERENCES 96

Tables and Figures 102

Supplementay Materials [tables are available at https://doi.org/10.57745/HRBVNR] 115

Pubblication: ***"A new flavi-like virus identified in populations of wild carrots"*** **119**

Pubblication: ***"Diversity of polerovirus-associated RNAs in the virome of wild and cultivated carrots"*** **122**

CHAPTER III **131**

Manuscript: ***"Metagenomic analysis of cultivated and wild carrots reveals different virome fingerprints and provides insight into viral fluxes between carrot populations"*** **132**

ABSTRACT 133

INTRODUCTION 134

MATERIALS AND METHODS 138

RESULTS 144

DISCUSSION 152

REFERENCES 161

Tables and Figures. 168

Supplementary Materials [tables are available at https://doi.org/10.57745/NAV8II] 178

GENERAL DISCUSSION **184**

1 Ecological and epidemiological questions and perspectives **185**

2 Etiological and agro-economic questions and perspectives **188**

BIBLIOGRAPHY **192**

ANNEX Publication *"Identification of two putative novel satellite RNAs with hammerhead ribozymes in the virome of French and Spanish carrot populatons"* **206**

GENERAL INTRODUCTION

1 Plant viruses – definition, classification and diversity

The plant virologist Richard Ellis Ford Matthews defined viruses as follows: "*A virus is a set of one or more nucleic acid template molecules, normally encased in a protective coat or coats of protein or lipoprotein, that is able to organize its own replication only within suitable host cells*" (Matthew, 1991). Therefore, viruses can be seen as biological entities that can parasitize all cellular life forms. Viruses differ from other pathogens in their size and shape, strategies of infection, multiplication and in the symptoms they produce. In particular, the genetic structure and replication of viruses differ from those seen in all cellular forms. The genomes of viruses vary significantly in size and can encode from one to about 250 proteins (Hull, 2014). Plant viruses possess rather small genomes but can reach sizes over 20kb, encoding up to 12 proteins (Candresse and Fuchs, 2020). The genetic material can consist of one or more nucleic acid molecules that can be packaged into separate particles as in the case of segmented viruses or into a single particle as for bi- or multipartite viruses. Another unique feature of viruses is their diversity of genome types and replication strategies. About 50 years ago David Baltimore classified viruses into six categories based on the types of genomes and the routes of genome expression, which are now known as Baltimore classes and cover the diversity of virus genomes and gene expression strategies. This represented an important conceptual framework in virology. The initial six classes comprise double stranded DNA (dsDNA), single stranded DNA (ssDNA), double stranded RNA (dsRNA), positive sense (+) ssRNA, negative sense (-) ssRNA and reverse transcribing (RT) RNA viruses (Baltimore, 1971). This was shortly after enriched by a seventh class comprising RT-dsDNA viruses that replicate through an RNA intermediate (Galibert *et al.* 1979). Further virus characteristics used for virus classification and nomenclature comprise the structure of virus particles and whether or not they are enveloped, host range, serological properties, pathogenicity and genome sequence similarities. At present, the International Committee on Taxonomy of viruses (ICTV) is the main governing body for all matters related to viral taxonomy. Until recently a 5-rank hierarchical system was used, comprising order, family (subfamily in some), genus and species. This was replaced by a 15-rank hierarchical system (Gorbalenya *et al.*, 2017) which is closer to the Linnaean taxonomic system used for cellular organisms. Another important change approved by the ICTV in 2021 was the adoption of a uniform binominal format for the naming of virus species. Comprehensive guidance on the application of these changes is provided by Sidell *et al.* (2020). With the advent of high-throughput sequencing, most viruses are today discovered through metagenomic studies

without easy access to their biological properties, and it has been proposed to reconsider the current rules for virus taxonomy to be able to take into account viruses identified solely on the basis of sequencing data from metagenomic studies (Simmonds *et al.*, 2017). Recent advances in metagenomic sequencing have shown that the actual diversity of viruses has been largely underestimated and that mixed infections are frequent and often symptomless in wild plants (Muthukumar *et al.*, 2009, Roossinck, 2012b, Roossinck *et al.*, 2015, Bernardo *et al.* 2018, Susi *et al.*, 2019). In contrast to the one host-one pathogen perspective, metagenomic studies have thus been able to explore virus communities in the context of different environments, which will be addressed in more detail below.

2 Plant virology - Historical perspective and methodological advances

The earliest record of a plant virus disease is possibly found in a poem composed by Japanese Empress Koken in 752 AD and translated by T. Inouye (Matthew, 1991):

"In this village
It looks as if frosting continuously
For the plant I saw
In the field of summer
The colour of the leaves were yellowing".

The poem has been suggested to describe the leaf yellowing symptoms caused in *Eupatorium lindleyanum* by a geminivirus, eupatorium yellow vein virus (Sauders *et al.*, 2003). Plant virology, and more generally virology as a branch of science, dates however back to the 1890s when the Russian botanist Dimitri Ivanovski and the Dutch chemist Martinus Beijerinck discovered tobacco mosaic virus and found that ultra-filtrates of infected plant extracts remained infectious (Fraenkel-Conrat, 1986). The fact that viruses were too small to be retained by bacterial filters lead to the development of more advanced techniques such as ultracentrifugation and electron microscopy (Brakke, 1953; Ball *et al.*, 1968; Baker *et al*., 1985) that provided more information about viruses shape and size. Especially in the second half of the 20th century much progress was made in the biological and molecular characterization of viruses by the development of a range of molecular tools such as protein and nucleic acids gel electrophoresis, serological techniques including enzyme-linked immunosorbent assay (ELISA) (Clark and Adams, 1977) or tissue blot immunosorbent assays (TBIA) (Lin *et al.*, 1990) and nucleic acids-based methods such as molecular hybridization techniques (Britten and Davidson, 1985) with the latter evolving into high-speed microarray assays specifically suited

for screening multiple viruses in a single diagnostic test (Schena *et al*., 1995; Boonham et al., 2007). Last but not least, the polymerase chain reaction (PCR) technique by which small amounts of DNA or RNA molecules could be amplified to a detectable amount of DNA (Henson and French, 1993; Hadidi and Candresse, 2003) was an important milestone in virus detection, characterization and diversity studies.

Another key technology, DNA sequencing, had a major impact on virology and, in general, on microbiology. Several approaches were developed but the most well-known is Sanger sequencing, developed in the mid 1970 (Sanger and Coulson, 1975) and referred to as first generation sequencing. Coupled with PCR technology, sequencing technologies revolutionized the view of microbial diversity as conserved genomic regions in bacteria and fungi, such as 16S and 18S ribosomal RNA genes could be amplified and sequenced, revealing many new taxons and the fact that the cultivable microorganisms represent just a minority of the actual diversity (Handelsman, 2005). In the last two decades, a new generation of sequencing technologies called Next Generation Sequencing (NGS) or High-Throughput Sequencing (HTS) has emerged, allowing the limitations of Sanger sequencing to be circumvented by the massively parallel sequencing of short reads (sequences), enabling whole genomes to be sequenced in a short time and generating enormous amounts of sequencing data. Different HTS technologies are available including second generation and third generation sequencing technologies (Heather and Chain, 2016). Second generation sequencing technologies (e.g. Roche 454, SOLiD or Ion Torrent/Life Technologies and Illumina/Solexa platforms) vary in their outputs, run times, and maximum read length and require different laboratory steps: (i) DNA fragmentation and the addition of synthetic DNA adapters (library preparation), (ii) clonal amplification of the prepared libraries (iii) massive parallel sequencing of DNA fragments (reads) (Massart *et al.*, 2014). Third generation sequencing (TGS), on the other hand, are innovative, long-read sequencing technologies (e.g. PacBio-Illumina or Oxford Nanopore Technologies (ONT)) capable of sequencing single molecules without the need to amplify DNA, as was the case with all previous technologies. TGS enables the generation of much longer reads (up to several kilobases) but is prone to higher error rates (Miclotte *et al.* 2016; Athanasopoulou *et al.*, 2022). However, the MinION instrument introduced by ONT is the only portable sequencing device that allows rapid real-time sequencing of long reads between 6000 and 48000 nucleotides (Deamer *et al.*, 2016), and has already found applications in plant virus diagnostics (Liefting *et al.* 2021).

2.1 HTS based metagenomics in plant virus diagnostics and ecology

Developments in HTS sequencing have driven the advent of the field of metagenomics, a term that first appeared in an article by Handelsman *et al.* (1998) and refers to the genetic study of all or nearly all genomes within an environmental sample (Handelsman, 2005; Roossinck *et al.*, 2015). Metagenomics have proven to be particularly interesting for virology as viruses lack universal genes such as 16S and 18S ribosomal genes present in bacteria and fungi. Indeed, the possibility of using shotgun sequencing (Breitbart *et al.*, 2002) or random primers for whole genome untargeted amplification (Roossinck *et al.*, 2010; Wang *et al.* 2002) has proven successful in circumventing the limitation due to the lack of these conserved genes in viruses. This has established the field of metagenomics of viruses (viromics). Several groups started to use virus metagenomic approaches in virus diagnostics of agricultural crops (Adams *et al.* 2009, Al Rwahnih et al, 2009; Kreuze *et al.*; 2009, Donaire *et al.*, 2009) and advanced in the development of efficient protocols and pipelines for virus detection. The use of HTS rapidly proved a compelling alternative to classical serological and molecular tests such as ELISA or PCR because it does not require prior knowledge of the pathogen(s) and allows to determine the complete phytosanitary status of a plant in a single assay (Studholme *et al.*, 2011; Maree *et al.*, 2018). In addition, the large amounts of virome sequence data obtained allow to differentiate virus variants that could contribute differentially to disease etiology (Maree *et al.*, 2018).

HTS-based metagenomics has greatly contributed to a new understanding of viral diversity in nature. One of the first plant virus metagenomic studies was performed in the frame of the "Plant Virus Biodiversity and Ecology" (PVDE) project that initiated a survey of virus diversity in native wild plants of the Tallgrass Prairie Preserve of north-eastern Oklahoma (Wren *et al.*, 2006; Melcher *et al.*, 2008.; Muthukumar *et al.* 2009). This precursor efforts began to uncover the largely underestimated viral diversity in natural ecosystems and gave rise to many other investigations on virus diversity in non-cultivated plants (Roossinck *et al.*, 2010, Lebouvier *et al.*, 2011; Roossinck et al., 2015; Maclot et al., 2020) as well as their role and distribution across agricultural and ecological boundaries (Malmstrom *et al.*, 2011; Alexander *et al.*, 2014; Bernardo *et al.*, 2018; Ma *et al.*, 2020, Maachi *et al.*, 2022). HTS has therefor revolutionized the field of plant virus ecology by closing the gap between classical virologists (focusing on disease-causing viruses in controlled and managed ecosystems) and ecologists (focusing on natural unmanaged ecosystems).

2.2 The plant virome and its technical challenges

Even though HTS-based metagenomics has already proven to be a powerful tool in plant virus diagnostics, ecology and epidemiology many technical challenges must still be met such as the development of efficient sample preparation methods or the choice of bioinformatic pipelines. Every HTS experiment starts with the sampling step that can be either single plants, pools of the same plants species, also referred to as the ecogenomics approach if the origin of these pools is georeferenced (Roossinck *et al.*, 2010) or complex pools involving different plant species representative of a certain area/environment. This last approach is sometimes referred to as the metagenomic or "lawnmower" method (Roossinck, 2012b). While the first strategy allows to keep host identity information and therefore allows some specific ecological or evolutionary analyses, the metagenomic approach allows for a more global virome comparison between environments or geographical locations. The next critical step is nucleic acid extraction, as field samples from plants contain complex mixtures of host and microbial (including viral) DNA and RNA. Different nucleic acid populations can be used for metagenomic studies in plants and have been described in detail (Roossinck *et al.*, 2012b; Roossinck *et al.*, 2015; Kutnjak *et al.*, 2021). In summary, the simplest method is to prepare total RNA or DNA from the sample of interest. This has the advantage of capturing the full spectrum of DNA and RNA viruses, reducing potential biases, but it inevitably results in data that are heavily skewed towards host plant sequences, complicating bioinformatic analyses and reducing virus detection sensitivity. Another approach targets small interfering RNAs (siRNAs), small 21-24 nt long RNAs produced by the host defence system as a result of viral RNA cleavage by the Dicer complex during viral infection. This method can be used to detect RNA or DNA viruses as well as viroids (Kreuze *et al.*, 2009; Donaire *et al.*, 2009), but the assembly of viral genomes can be challenging due to the small size of siRNAs reads. When it comes to more complex plant pools, the enrichment of viral nucleic acids may allow to capture also low titre viruses. Two enrichment approaches are available: (i) purification of dsRNAs, a molecule that is a hallmark for RNA viruses and virus-like agents' replication in plants (Dodds *et al.*, 1984) but has the disadvantage that DNA viruses may be counter selected and (ii) virion associated nucleic acids (VANA) which is probably the most widely used virus enrichment approach in metagenomic studies (Moubset *et al.*, 2022). The VANA approach relies on filtration, ultracentrifugation and nuclease treatment steps to enrich both RNA and DNA virus particles but the method may counter-select viruses with labile particles or particle-less agents.

The computational analysis of huge sequencing datasets generated also represents a very significant challenge in HTS based metagenomics. Most published virome studies follow a common series of steps: (i) removal of low-quality reads and trimming of adapter sequences (ii) *de novo* assembly of reads into contigs (contiguous nucleotide sequences assembled from overlapping reads) and (iii) identification of viral reads or contigs against a database of known viral sequences or motifs. The most widely used databases are the NCBI non-redundant nucleotide database (nr/nt,), the non-redundant GenBank protein database (nr) or the viral RefSeq database used for BLAST approaches (Altschul *et al.*, 1990) or DIAMOND (Buchfink *et al.*, 2014). Alternatively, viral protein motifs databases such as PFAM (Punta *et al.*, 2012) can be used, together with specific, motif-searching algorithms that have been shown to be faster (Kutnjak *et al.*, 2021). If the main scope is to identify known viruses, quality-filtered reads can also be directly mapped against reference viral genomes. Virus identification can be performed using either reads or *de novo* assembled contigs. However, longer sequences obtained through an assembly step generally improve the ability to identify viruses and to reduce the volume of unannotated "dark matter" (François *et al.*, 2018). In an ideal situation, *de novo* assemblers can reconstruct near complete viral genomes from sequencing reads with a uniform read coverage along the genome. Examples of frequently used assemblers in viral metagenomic studies comprise e.g. Velvet, IDBA-UI, Spades and MetaSpades or assemblers present in more user-friendly, commercially available packages such as CLC Genomics Workbench or Geneious prime (https://www.geneious.com/) which are reviewed in detail elsewhere (Blawid *et al.*, 2017; Kutnjak *et al.*, 2021). The assembly step faces several challenges, including low DNA yields limiting read coverage that can lead to fragmentation in the assembly and to a decrease in the recovery of the genomes (Hasiów-Jaroszewska *et al.*, 2021). Another issue that has to be addressed in HTS experiments is contamination (Massart *et al.*, 2014; Galan *et al.* 2016), that can be introduced during all wet-lab steps necessary for library preparation but also when multiple libraries are sequences in a multiplexed format which can lead to index hopping (Ilumina, 2017; van der Valk *et al.*, 2019). Specific precautions and the use of positive and negative controls or of the recently proposed alien controls is highly recommended to reduce or eliminate contamination (reviewed by Massart *et al.*, 2022).

In summary, the choice between sample preparation methods, sequencing platforms and bioinformatic pipelines can require careful strategic choices that take into account the study specific objective(s) as well as the budget and know-how available. Careful planning is also required when it comes to controls and standards, as discussed in more detail in Chapter 1.

3 Plant virus ecology and evolution

Although early research on viruses was more ecological in nature (Broadbent, 1950; Bos, 1981; Gibbs, 1983), the introduction of molecular tools led to a more restricted view, as viruses were mainly studied on a molecular basis in terms of their potentially pathogenic character in systems of anthropological importance, be it as human pathogens or as agricultural pests.

Plant virus ecology considers viruses in the light of entire environments including biotic and abiotic factors. It draws from different disciplines including virology, plant biology and ecology, epidemiology, and entomology (Hull, 2014). Plant viruses are therefore studied in a range of environments to assess their diversity, interaction with vectors or with other viruses in mixed infections and their contribution to the functioning of plant populations and communities in more diverse systems such as natural, non-cultivated ecosystems. Earlier studies of viruses in wild plants have shown that virus-induced changes are not so common and that infections are frequently symptomless (Cooper and Jones, 2006). In this context, wild plants have been found to be frequently infected with persistent viruses (Roossinck, 2012a, Roossinck, 2015a; Roossinck 2015b) that do not cause visible symptoms. Persistent plant viruses, also called "cryptic" or latent viruses were first described in the 1960s (reviewed in Boccardo *et al.* 1987). These viruses are only transmitted vertically through seeds or pollen (Valverde and Gutierrez, 2007), and appear to infect their hosts for many generations. In contrast to viruses transmitted horizontally (mechanically or through vectors), persistent viruses lack a gene for systemic movement. They rather move through their host cell divisions, thereby infecting every host cell (Roossinck, 2015a). Many families of persistent viruses are shared between plants and fungi and have mostly dsRNA genomes, such as the *Amalgaviridae*, *Chrysoviridae*, *Partitiviridae* and *Totiviridae*. In this context, the *Endornaviridae*, classified with ssRNA genome viruses, are an exception. Since they are very generally asymptomatic, there is little information about the impact of persistent viruses on their hosts but evolutionary arguments and a few studies suggest that some persistent viruses might have a positive impact on their host plants. It has for example been shown that the coat protein of white clover cryptic virus 1 suppresses nodulation of legumes if adequate soil nitrogen is available (Nakatsukasa-Akune *et al.*, 2005). Virome studies, and in particular those using a dsRNA-based HTS approach, have demonstrated that persistent viruses are abundant in wild plants where they may represent more than half of the viral sequences that show homologies with entries in GenBank (Roossinck *et al.*, 2010, Roossinck, 2015, Roossinck and Garcia-Arenal, 2015, Fetters *et al.* 2022). These and similar findings have shifted our view of viruses as purely disease-causing agents and highlighted the

need to get a deeper insight into their *raison d'être*. We might only be at the beginning of the elucidation of the contribution of these biological entities to the functioning of ecosystems and of the importance of the 'ecological trinity' between virus, vector and host.

3.1 The ecological trinity

3.1.1 Virus vector interactions

Other than vertical transmission through seeds or pollen described earlier, some viruses are transmitted effectively by vectors. Most vector-borne plant viruses are transmitted by insect vectors of the order Hemiptera, which include aphids, whiteflies, leafhoppers, and thrips (Whitfield *et al.*, 2015), but transmission by mites, soil-borne nematodes, chytrids, and plasmodiophorids also occurs (Roossinck, 2015b). Hemipteran insects are especially efficient vectors because of their needle-like stylets with which they feed on plants. Depending on transmission parameters (acquisition period, latency period, retention period during which the vector remains viruliferous), several transmission modes have been identified, from non-persistent (retention period of minutes to hours), semipersistent (retention up to a few days) and persistent (life-long retention). Transmission mode also somewhat correlates with virus localization in the vector: non-persistent viruses are stylet-borne, semipersistent viruses are foregut- and/or stylet borne while in persistent transmission the virus circulates in the entire insect body and can be either non-replicative or replicative, depending on whether it propagates/replicates within the vector. The persistent transmission mode is also characterized by lengthy acquisition and latency periods. In contrast, for non-persistently transmitted viruses, a brief probing of infected hosts is sufficient to acquire virions and for the vector to become viruliferous. These transmission modes have been extensively reviewed elsewhere (Hogenhout et al, 2008; Whitefield *et al.*, 2015). It should be emphasized that the interaction between viruses and vectors may contribute to the determination of the realized natural host range of viruses. For example, carrot red leaf virus host range is limited to the *Apiaceae* family (Watson and Falk, 1994; Yoshida, 2020) and it is naturally only transmitted by the willow carrot aphid *Cavariella aegopodii*, in a persistent non-propagative way. However, under laboratory conditions, it was also transmitted by the very polyphagous aphid *Myzus persicae* (Naseem *et al.*, 2016).

3.1.2 Virus-host interactions

Virus infection can affect the host plant at different levels, (i) at the cellular level and (ii) at the broader ecological level. At the cellular level, plants usually respond to virus infection with various defence mechanisms such as RNA interference (RNAi) also termed RNA silencing. The replicative dsRNA forms of viruses are recognized the Dicer complex and cleaved into small 21-24-nt RNAs. As countermeasures, viruses have evolved RNA silencing suppressors, suggesting a long coevolution of plant hosts and viruses (Csorba *et al.*, 2015). While virus-infected plants may exhibit various abnormalities such as leaf mosaic patterns, necrosis, chlorosis, stunting, etc. (Hull, 2014), the final outcome of cellular interactions between viruses and plants may also depend on abiotic factors, such that the interaction between virus and host can change from harmful to beneficial under certain environmental conditions (Fraile and García-Arenal, 2016). For example, some reports have shown that under abiotic stress conditions viruses enhance stress tolerance in plants (Xu *et al.*, 2008; Gorovits *et al.*, 2019). Viruses may also interact with other co-infecting viruses within host plants. Mixed infections are frequent and co-infecting viruses can show interactions over a wide spectrum ranging from synergistic to antagonistic. Synergistic interactions refer to a benefit for at least one of the viruses involved manifested, for example, by increased replication and/or pathogenicity resulting in more severe symptoms than those observed in single infections. A well characterized example is the synergism between potato virus Y (PVY, genus *Potyvirus*) and potato virus X (PVX, genus *Potexvirus*) characterized by a 3- to 10-fold increase in PVX titer as compared with single infection (Pruss *et al.*, 1997). Conversely, when one virus causes a decrease in replication or pathogenicity of the other virus, the interaction is said to be antagonistic. Antagonism plays for example an important role in cross-protection, otherwise referred to as superinfection exclusion (SIE) (reviewed by Mascia and Gallitelli, 2016).

Another type of interaction is termed assistance, where a dependent virus, defective in one or more functions, can see these functions restored through the action of a ‘helper’ virus. Helper-dependence is exemplified by the interaction between umbraviruses (which lack a coat protein gene) and viruses of the family *Solemoviridae* (previously *Luteoviridae*) that are able to encapsidate umbraviruses genomes in their own particles, in a process called transcapsidation. Transcapsidated umbraviruses can then benefit of the aphid transmission possibilities offered by *Solemoviridae* particles. Conversely, umbraviruses may facilitate the *in planta* movement of their helper viruses through the action of their movement protein. Well characterized examples of such interactions include groundnut rosette virus (GRV, Umbravirus) and its dependency on

Groundnut rosette assistor virus (GRAV) (Robinson *et al.*, 1999) or viruses involved in the carrot motley dwarf disease (CMD) (Watson and Falk,1994; Yoshida, 2020).

At the ecosystem level, viruses can influence host fitness with respect to competitors in complex host communities. Metagenomic studies have shown that different plant hosts are usually infected with multiple viruses and are rarely associated with disease, especially in natural plant populations (Roossinck, 2012b; Roossinck 2015b, Thapa *et al.*, 2015, Susi *et al.*, 2019; Ma *et al.*, 2021). However, the absence of symptoms does not necessarily mean that viruses have no impact on their hosts. Malmstrom *et al.* (2005) observed that introduced exotic grasses in California exhibited higher aphid infestations and resulted in increased virus incidence in native perennial grasses, most likely reducing their fitness in the face of environmental stress and thus altering plant population dynamics to the benefit of the introduced grasses.

3.1.3 Virus-host-vector interactions

Virus ecology is intricately linked to the ecology of both insect vectors and host plants. Plant viruses are known to impact the behaviour of insect vectors directly or indirectly, influencing the frequency and nature of virus-vector interactions and therefore the frequency of transmission events. In a broad literature analysis, Mauck *et al.* (2012) addressed the hypothesis that different modes of vector transmission have different effects on host traits that mediate vector attraction, arrestment on and dispersal from infected plants. The authors found evidence that virus-induced changes in host characteristics generally favoured vector attraction, regardless of the mode of transmission. However, non-persistently transmitted viruses often reduce plant quality to ensure short feeding times and rapid spread, while persistently transmitted viruses tend to improve host quality to promote long-term feeding. Accordingly, CMV-infected plants exhibited increased emissions of plant volatile compounds, which are known to be a key foraging stimulus for aphids, increasing attractiveness to vectors. At the same time, host plant quality decreased as a consequence of the production of anti-feeding compounds, prompting aphids to quickly move away and migrate to new plants, a strategy that promotes transmission of CMV, which requires very short feeding periods for both acquisition and transmission (Carmo-Sousa *et al.*, 2014; Mauck *et al.*, 2010). In addition, CMV-infected tomato and Arabidopsis have been shown to alter the foraging behaviour of bumblebees (*Bombus terrestris*), thereby increasing pollination and plant reproduction (Groen *et al.*, 2016). The authors speculate that this might be a ‘payback’ of the virus to compensate for a decreased yield of seeds observed in CMV infected plants in the absence of pollinators, and that such

enhanced pollinator services might provide mutual benefits to the virus and to its susceptible hosts in wild plant populations.

3.2 Viruses at the agroecological interface

Human activities such as agricultural practices and increasing levels of globalization have caused heavily managed agroecosystems to be more and more frequently located adjacent to natural undisturbed habitats, creating what has been termed the agroecological interface. These boundaries represent an interactive area populated by crop plants, weeds, and wild non-crop plants as well as diverse animals and microbes (Alexander *et al.*, 2014) A this boundary, plant genes can be exchanged between crop plants and wild relatives via the movement of pollen and seeds (Alexander *et al.*, 2014). Microbes and in particular viruses, as highly adaptive and variable biological entities, are equally able to move in both directions through the agroecological interface, from wild plants to crops and vice versa. Early studies of wild plants have indeed focused on their role as a source of infection for crop plants as “reservoir hosts”, ensuring the maintenance of viruses when host crops are absent from the field e.g. during winter or during crop rotations (Watson and Serjeant 1964; Bos, 1981). The paucity of viral diseases in wild plant communities has been ascribed to the long coevolution between wild plants and viruses (Lovisolo *et al.*, 2003) or to other ecological factors limiting for viral populations, such as lower host plant densities, the presence of predators that limit vector abundance of vectors or the mixture of different genotypes and non-host plants that "dilute" virus spread (Cooper and Jones, 2006). Many crop viruses are believed to be originated from wild relatives that have contributed to the development of crop species. The loss of biodiversity, the increased prominence of the agroecological interface, together with enhanced global trade are contributing to an increase in "new encounters" between otherwise separate plant and virus species and have likely increased the frequency of virus emergence events (Roossinck and Garcia-Arenal, 2015). Emergence can be envisionned as involving several sequential steps: (i) virus jump from a reservoir host to a new host species (or to the same host species in different environmental or ecological conditions), (ii) virus adaptation to the new host or environment, (iii) spread in the new host population through enhanced between-host transmission, often as a consequence of adaptation to vectors. These phases are extensively reviewed by Elena *et al.* (2014). Emergence often goes along with changes in pathogenesis and enhanced disease severity, leading to devastating epidemics (Fargette *et al.*, 2006; Elena *et al.*, 2014; Roossinck and Garcia-Arenal, 2015). The subtropics and tropics provide many examples where viruses emerged from indigenous plants at the agroecological interface such as cassava mosaic virus

(family *Geminiviridae*, genus *Begomovirus*) causing serious epidemics in cassava after its introduction to Africa from South America (Legg, 1999). Advances in HTS methods provide the ability to analyse virus communities in wild and adjacent crop species in comparative studies and to identify potential virus reservoirs (Roossinck *et al.* 2015; Ma *et al.*, 2020; Maachi *et al.*, 2022). These studies show that the virome of wild plants can significantly differ from that of neighbouring cultivated plants. In fact, a poorly investigated possibility it that newly introduced crop viruses may spill over from cultivated plants to wild plant communities, with the potential to threaten biodiversity. Most of the sparsely available data suggest that the impact of crop viruses on fitness components of wild plants, in terms of survivalship or reproductive success, is largely negative or neutral (Maskell *et al.*, 1999; Malmstrom *et al.*, 2005; Malmstrom and Alexander, 2016).

Infection risk may therefore depend on host and vector ecology with ecosystem diversity likely to be a key determinant. Two opposing hypotheses have emerged. First, it is hypothesized that increased plant diversity may lead to an increased risk of infection due to higher abundance of host plants and reservoirs, is what has been termed the “amplification effect”. On the other hand, increased diversity may also lead to a reduction in the abundance of the most competent hosts in what is referred to as the dilution effect. Evidence has been reported in support of both hypotheses but the majority of studies to date found evidence for the dilution effect and thus a negative correlation between biodiversity and disease risk (Pagán *et al.*, 2012, Civitello *et al.*, 2015, Fraile *et al.*, 2017, Liu *et al.*, 2020).

4 Carrots and their viruses

4.1 Biology and life cycles of *Daucus carota* L.

The family *Apiaceae* comprises about 3,700 species consisting of wild plants and important crops (Petruzzello, 2019). Carrot (*Daucus carota* L) is the most important root vegetable of the family and is cultivated worldwide (Que *et al.*, 2019). There is general consensus with evidence from genetic studies that the domesticated carrot likely developed from the wild carrot in Central Asia, around 1,100 years ago (Iorizzo *et al.*, 2013, Rong *et al.*, 2014). Carrot is a diploid, highly outcrossing and insect pollinated species. Plants are andromonoecious (they possess bisexual flowers and are therefore self-pollinating, but can also be cross-pollinated). Some populations are found to exhibit male sterility (Ronfort *et al.*, 1995), a trait widely used in the development of cultivated hybrid carrot. Flowers are organized as umbels, that consist of many tiny flowers held on short flower stalks. Carrots are cold-tolerant plants, which is why they are

also classified as winter vegetables. Vernalization (periods of low temperatures between 0 and 10 °C following the seedling stage) is required for flowering. Carrot is a biennial plant and during the first year it produces a basal rosette of leaves and stores carbohydrades and sugars in a tuberized tap root. After the development of a substantial number of leaves (8-12) and when storage roots have reached 4-8 mm in diameter plants become responsive to vernalization, though the duration of the required cold period varies greatly across genotypes (Linke et al., 2019). After a vernalization period, flowering starts normally during the spring and summer months of the following year. The emergence of flower stems is called bolting. Once reproduction is completed and a plant has set seed it will die (Mandel *et al.*, 2019). Wild carrots (*Daucus carota* ssp *carota*) however might not always act as biennials and show some variation in their life history strategies. Wild carrots may behave as winter annuals, germinating in fall and surviving winter as rosettes and flower the next summer. They can also occur as perennials, surviving several years as rosettes with postponed flowering. These different life cycle strategies may be highly dependent on genetic and environmental factors with, for example, annual plants being more common in resource-rich environments and perennials in nutrient-poor environments (Wijnheijmer, 1989; Mandel *et al.*, 2019). In the carrot industry, the roots are either harvested directly for consumption in the first season or, if grown for seed production, stored at low temperatures to protect them from winter frosts while still providing the cool needed for vernalization. The roots are then planted in spring and produce seeds that are harversted around October. Wild and cultivated carrots belong to the same species and can easily hybridize. A common phenomenon in the cultivation of carrots is the blooming of some carrot plants in the field during the first year (Figure 1). These plants are also called bolters and have woody whitish roots, a characteristic also observed in wild carrots. It is thus generally assumed that these bolters result from hybridization with cultivated carrots due to contamination during seed production. The early bolting in the field is likely related to the fact that wild carrots, especially in Europe, often have a one-year life cycle, unlike biennial cultivated carrots (Wijnheijmer *et al.*, 1989, Mandel *et al.*, 2019).

These so-called off-type carrots represent a problem for carrot producers as they are of no economic value. The main carrot seed production areas in the northern hemisphere are located in the northwestern USA (Oregon, Washington, Idaho and northern California) and in southern France (Spurr and Lucas., 2021). France is one of the top producers of carrots in Europe and the region of Nouvelle Aquitaine represents the largest carrot growing area. Most of the regional production (99%) is located in the departments of Gironde and Landes (draaf.nouvelle-

aquitaine.agriculture.gouv.fr), where deep sandy soils provide optimal growing conditions. Wild carrots are common in temperate regions, frequently found in undisturbed sites, along roads or abandoned fields. In Gironde and Landes, it is the only species within the *Apiaceae* family significantly growing in deep sandy soils outside of wetter areas along streams.

Figure 1. Bolter carrots growing within cultivated carrot fields in Nouvelle Aquitaine (France). These plants with phenotypic traits of wild carrots (early bolting, small tap roots) are considered to be hybrids between wild and cultivated carrots contaminating commercial seed lots.

4.2 Carrot viruses

In addition to various bacterial and fungal pathogens, viruses are among the most important pathogens of carrots. In particular, aphid-transmitted viruses with +ssRNA genomes of the families *Tombusviridae*, *Solemoviridae*, *Closteroviridae*, *Secoviridae* and *Potyviridae* include well-known carrot-infecting viruses that can cause severe damage to carrot production. A list of the most important carrot infecting viruses with corresponding literature references is given in Table 1.

Table 1: known viruses infecting carrots (*Daucus carota* L.)

Virus family	Virus genus	Virus name	Category	References
Solemoviridae	*Polerovirus*	carrot red leaf virus	(+)ssRNA	Watson et al., 1964, Yoshida, 2020
Solemoviridae	*Polerovirus*	wild carrot red leaf virus		
Tombusviridae	*Umbravirus*	carrot mottle virus		Watson et al., 1964, Yoshida, 2020
Tombusviridae	*Umbravirus*	wild carrot mottle virus		
Tombusviridae	*Umbravirus*	carrot mottle mimic viurs		Gibbs et al., 1996
Closteroviridae	*Closterovirus*	carrot yellow leaf virus		Adams et al., 2014
Closteroviridae	*Closterovirus*	carrot closterovirus		Adams et al., 2014
Closteroviridae	*Crinivirus*	lettuce infectious yellows virus		Duffus et al., 1986
Betaflexiviridae	*Chordovirus*	carrot Ch virus 1		Adams et al., 2014
Betaflexiviridae	*Chordovirus*	carrot Ch virus 2		Adams et al., 2014
Betaflexiviridae	*Carlavirus*	carrot carlavirus WM-2008		Menzel and Vetten, 2008a
Potyviridae	*Potyvirus*	carrot virus Y		Moran et al., 2002
Potyviridae	*Potyvirus*	carrot thin leaf virus		Howell and Mink, 1976; Howell and Mink, 1981
Potyviridae	*Potyvirus*	*celery mosaic virus*		Howell and Mink, 1981
Potyviridae	*Potyvirus*	*watermelon mosaic virus*		Parry and Persley, 2005
Potyviridae	*Potyvirus*	gazar virus Y		Soliman et al., 2012
Secoviridae	*Sequivirus*	carrot necrotic dieback virus		Menzel and Vetten, 2008b
Secoviridae	*Sequivirus*	parsnip yellow fleck virus		van Dijk and Bos, 1985
Secoviridae	*Torradovirus*	carrot torradovirus 1		Adams et al,, 2014; Rozado-Aguirre et al., 2016
Bromoviridae	*Alphamovirus*	alfafa mosaic virus		Howell and Mink, 1981
Bromoviridae	*Cucumovirus*	cucumber mosaic virus		Afreen et al., 2009
Pospiviroidae	*Pospiviroid*	citrus exocortis viroid	circular (-)ssRNA	Fagoaga and Duran-Vila, 1996
Partitiviridae	*Alphapartitivirus*	carrot cryptic virus	dsRNA	Willenborg et al., 2009
Partitiviridae	*unassigned*	Carrot temperate virus 1		Natsuaki et al., 1990
Partitiviridae	*unassigned*	Carrot temperate virus 2		Natsuaki et al., 1990
Partitiviridae	*unassigned*	Carrot temperate virus 3		Natsuaki et al., 1990
Partitiviridae	*unassigned*	Carrot temperate virus 4		Natsuaki et al., 1990
Geminiviridae	*Begomovirus*	tomato leaf curl New Delhi virus	dsDNA	Sivalingam et al., 2011
Geminiviridae	*Begomovirus*	pedilanthus leaf curl virus		Saritha et al., 2017
Geminiviridae	*Begomovirus*	ageratum enation virus		Kumar et al. 2013
Tolecusatellitidae	*Betasatellite*	ageratum yellow leaf curl betasatellite	circular ssDNA	Kumar et al. 2013
unclassified	*unclassified*	carrot mottle virus satellite RNA	linear ss satellite RNA	Menzel et al., 2009
unclassified	*unclassified*	carrot red leaf virus associated RNA	Tombus-like associated RNA	Watson et al., 1998
unclassified	*unclassified*	beet western yellows virus associated RNA		Adams et al., 2014

One of the first identified virus disease in carrots is the carrot motley dwarf (CMD) disease first described by Stubbs in Australia (1948). CMD is caused by a complex interplay between a polerovirus, carrot red leaf virus (CtRLV), an umbravirus, carrot mottle virus (CMoV) or carrot mottle mimic virus (CMoMV) and, in some instances, a polerovirus-associated RNA, carrot red leaf virus-associated RNA (CtRLVaRNA) that interact in a helper-dependent way (Yoshida, 2020). The umbravirus and the associated RNA lack a capsid protein and are transcapsidated by the coinfecting polerovirus, thus allowing their transmission by the willow carrot aphid (*Cavariella aegopodii*). Other early studies have identified carrot necrotic dieback virus (CNDBV, genus *Sequivirus*). CNDBV, originally described as the Anthriscus strain of parsnip yellow fleck virus (PYFV, genus *Sequivirus*) by van Dijk and Bos (1985) naturally infects cow parsley (*Anthriscus sylvestris* L.) but causes leaf necrosis, severe stunting and dieback in carrots (Menzel and Vetten, 2008b). Both, PYFV and CNDBV are transmitted by the aphid *Cavariella aegopodii* in a semipersistent way but transmission occurs only in case of coinfection with the helper virus Anthriscus yellows virus (AYV, genus *Waikavirus*, family *Secoviridae*). *Cavariella aegopodii* exhibits seasonal host alternation with *Salix* spp. as primary hosts, where sexual generations and overwintering occurs and a secondary more nutritious herbaceous host of the family *Apiaceae*, where parthenogenetic reproduction occurs (Kundu and Dixon, 1993, Abbot *et al.*, 2018). In spring the eggs hatch on the primary host and winged migrants are formed which migrate to the secondary herbaceous host with a peak at the end of spring (Dunn, 1965). On the secondary host, the aphids produce several asexual parthenogenetic cycles of winged and wingless morphs throughout summer until in fall winged males and gynoparae (winged morphs giving birth to sexual females) are formed. Gynoparae migrate back to the primary willow hosts to produce sexual females (oviparae) which then mate with the males and produce overwintering eggs (Kundu and Dixon, 1993; Williams and Dixon, 2007). The life cycle of *Cavariella aegopodii* is represented in figure 2. As pointed out by early studies, overwintering wild plants of the *Apiaceae* family may play an important role as source of virus inoculum, and therefore in the spread of carrot motley dwarf to carrot crops (Dunn, 1965, Watson and Serjeant, 1964).

In addition to viruses involved in complex assistance interplays, other ssRNA viruses are known to infect carrot crops. Carrot thin leaf virus, a potyvirus, is a regularly occurring pathogen of commercial carrot crops in western USA (Geoffriau and Simon, 2020) and first reported from field carrots in Washington State. It causes vein clearing and leaf narrowing on carrot seedlings (Howell and Mink 1976).

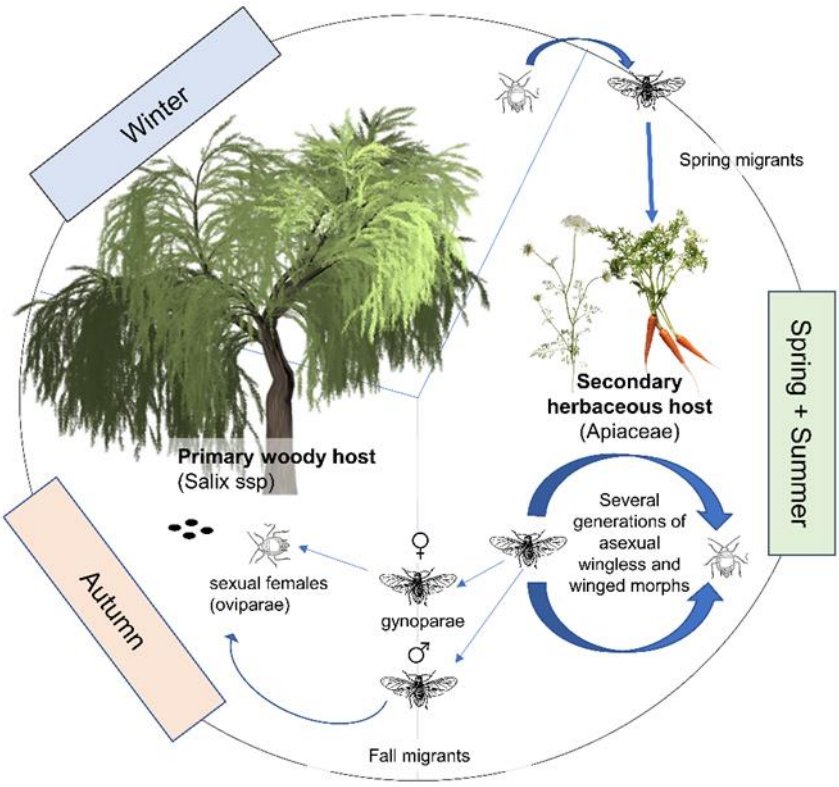


Figure 1: Life cycle of *Cavariella aegopodii*

Another Potyvirus affecting carrot crops is carrot virus Y, which causes foliar and root disease and results in significant yield reduction. Its occurrence has been limited to the Australian continent (Moran *et al.*, 2002; Geoffriau and Simon, 2020). Some viruses from the family *Closteroviridae* can also affect carrot crops, including carrot yellow leaf virus (CYLV) or carrot closterovirus, with the former being associated with internal root necrosis (Adams *et al.* 2014) and the later having been recently discovered by HTS-based virome scanning. Through the same HTS analyses, the authors also identified two novel viruses of the genus *Chordovirus*.

Up to date, only a few viruses with dsRNA genomes have been identified in carrots, including a member of the *Partitiviridae* family, carrot cryptic virus (CaCV, genus *Alphapartitivirus*) reported in Germany (Willenborg *et al.*, 2009) and several unassigned *Partitiviridae* members reported in Japan (Natsuaki *et al.*, 1990) and referred to as carrot temperate viruses 1-4 (CteV1-4).

4.3 Carrot as model plant for virome comparisons

In efforts to determine the effects of intrinsic host traits on host range Moury *et al.* (2017) found that the plant family taxonomic rank represents an important threshold for virus host range, indicating that ability for a virus to infect a plant species is correlated with the botanical closeness of this species to hosts of the virus. In addition, the host genotype has been shown to play a key role in the composition of virus communities (Sallinen et al, 2020). Cultivated and wild carrots belong to the same species and therefore represent the lowest genetic and botanical barrier to virus flow between a crop and its wild relative. In the present work, wild and cultivated carrot populations have been used as model pathosystem to compare their viromes and to analyse virus fluxes between wild and cultivated compartments.

5 Outline of the scientific tasks and questions addressed in the present thesis

In the work presented here I used HTS-based virus metagenomics approaches to address both methodological and virological/ecological questions related to plant virome description and comparison.

The methodological part, presented in **Chapter 1**, is based on the development and use of phytovirus synthetic communities with different degrees of complexity to benchmark the performance of VANA and dsRNA-based HTS approaches. Specifically, the following questions have been addressed: (i) how accurate and complete is the description of the virome we obtain? (ii) How is virome description affected by different HTS strategies (VANA, dsRNA)? (iii) what other parameters influence virome description performance? The results are presented in Chapter 1 as a manuscript that will be submitted shortly.

With regard to virological questions, I have used a dsRNA-based metagenomic approach to study the virome of 45 populations of different wild and cultivated carrots in five different regions of France, as well as six additional populations from central Spain obtained from Spanish collaborators, Prof. Fernando Garcia Arenal and his PhD student Bisola Babalola. The results are presented in **Chapter 2** and provide a full description of the identified virome. It is structured in the form of a manuscript (to be submitted shortly) describing all novel and known viruses detected in the different carrot populations, and of two recently published articles, one describing a novel flavi-like virus in wild carrots and the other discussing the diversity of polerovirus-associated RNAs identified in the carrot virome.

In the **Annex**, a recently published paper, on which I am a co-author, is included. It describes satellite RNAs with hammerhead ribozymes identified in the virome of French and Spanish carrot populations.

The third part, presented in **Chapter 3** as manuscript to be submitted shortly, deals with more ecological aspects, including a detailed virome analysis to compare virus community fingerprints between cultivated and wild carrots at the local level in Nouvelle-Aquitaine, France. A complementary approach of HTS and RT-PCR has allowed to extend the analysis from the plant population level to individual plants, and to move from the virus interspecies level to the analysis of virus populations to gain insights into virus fluxes (or barriers to fluxes) at the agroecological interface. Therefore, the following questions could be addressed: (i) how different are the viromes in cultivated and wild carrot populations? (ii) can virus exchanges (or absence of exchanges) between wild and cultivated carrot populations be identified? (iii) can non-random virus interactions within and between host types be identified?

CHAPTER I

Benchmarking of virome metagenomic analysis approaches using a large, 60+ members, viral synthetic community

Benchmarking of virome metagenomic analysis approaches using a large, 60+ members, viral synthetic community

D. Schönegger[1], O. Moubset[2,3], P. Margaria[4], W. Menzel[4], S. Winter[4], P. Roumagnac[2,3], Armelle Marais[1] and Thierry Candresse[1*]

[1]Univ. Bordeaux, INRAE, UMR BFP, CS20032, 33882, Villenave d'Ornon Cedex, France.

[2] *CIRAD, UMR PHIM, 34090 Montpellier, France*

[3] *PHIM Plant Health Institute, Univ Montpellier, CIRAD, INRAE, Institut Agro, IRD, Montpellier, France*

[4] *Leibniz-Institute DSMZ, 38124 Braunschweig, Germany*

*** Author for correspondence:** thierry.candresse@inrae.fr Tel : (+33) 557 12 23 89

Abstract word count: 231 words

Text word count: 6117 words (Introduction-Materials and Methods-Results-Discussion)

Tables and Figures: 5 Tables, 4 Figures, 7 supplementary Tables and Figures

Running title: Synthetic viral community for virome benchmarking

Abstract

In contrast to other areas of metagenomics, while a range of approaches and target nucleic acid populations have been used for the metagenomic analysis of viromes, there has to date been only limited efforts to benchmark their performance, in terms both of completeness of description and of faithfulness to community structure. While natural communities are more readily accessible, synthetic communities constructed using well characterized isolates present added advantages since they allow to evaluate performance in absolute terms. Starting from well characterized, quality controlled viral isolates from the DSMZ collection, we have assembled synthetic communities of varying complexity up to a highly complex community of 68 viral agents (115 viral molecules) comprising isolates from 21 viral families and 63 genera. These communities were then analyzed using a virion-associated nucleic acids (VANA) based strategy and a highly purified double-stranded RNAs (dsRNA) based one. The results obtained allow to compare the performance of these two approaches for groups of viruses and satellites with different genome types and confirm that the dsRNA-based approach provides a more complete representation of the RNA virome. They also allow to identify important parameters and to propose hypotheses to explain differences in performance, in particular the imbalance in representation of individual viruses using each approach. Remarkably, these analyses highlight a strong relationship between the completeness of virome description and sample sequencing depth which should prove useful in virome analysis efforts.

Keywords: virome, VANA, dsRNA, synthetic community, metagenome, double-stranded RNA, high-throughput sequencing

INTRODUCTION

Significant advances in the development of molecular methods have been made in the last decades, including innovative sequencing technologies based on DNA/RNA approaches such as targeted (RT)-PCR or non-targeted High-Throughput Sequencing (HTS). HTS, also known as next generation sequencing (NGS), enables high-speed, high-throughput sequencing of the entire genetic material (metagenome) in a sample, generating enormous amounts of sequencing data. These developments led to major advances in the field of metagenomics, i.e. the sequencing of the entire genetic material of a sample, and to a new understanding of the diversity of the microbial world (Zhang *et al*. 2019; Jian *et al*., 2021). Viral metagenomics has revealed the immense diversity and ubiquity of viruses in nature and thus revolutionized our vision of these biological agents (Lefeuvre *et al*., 2019, Maclot *et al*., 2020; Roux *et al*., 2021). Specifically, these metagenomics studies have revealed that sequence data available in public databases are biased toward human viruses or viruses of anthropological significance, with e.g. influenza-like viruses found in fish and amphibian hosts (Shi *et al.*, 2018) or more than 75% of the characterized plant virus species that have been isolated from crops (Wren *et al.*, 2006). These, together with findings of viruses associated with hosts different from those known for the vast majority of their relatives such as flavi-like viruses found in plants (Kobayashi *et al*., 2013, Schönegger *et al.*, 2022) have raised novel questions about virus-hosts co-divergence or host switching. In plant virology in particular, advances in the development of viral metagenome analyses have been of great importance, as more than half of emerging diseases in plants are thought to be caused by viruses (Anderson *et al.* 2004), while early detection of causative agents is critical to prevent or control disease outbreaks. Plants are very frequently infected by more than one virus (Moreno and López-Moya, 2020), complicating the unravelling of the etiology of plant viral diseases. HTS has a huge potential in plant virus diagnostics because it allows to picture the complete phytosanitary status of a plant and to differentiate

between virus variants that may contribute differentially to disease etiology (Maree *et al.*, 2018). For example, in a metagenomic analysis of sour cherry showing symptoms of Shirofugen stunt disease (SSD), a divergent isolate of little cherry virus 1 (LChV1) was identified in the absence of any other viral agent, suggesting that LChV1 could be responsible for the SSD disease (Candresse *et al.* 2013).

HTS has also renewed the link between classical plant virology and ecology (Malmstrom *et al.*, 2011; Maclot *et al.*, 2020). The rich viromes identified in different plants and plant populations enabled the study of ecological processes such as the movement of viruses between different host reservoirs, the effects of management practices or of the anthropological simplification of ecosystems (Bernardo *et al.*, 2018, Ma *et al.*, 2020, Ma *et al.*, 2021; Susi and Laine, 2021; Maachi *et al.*, 2022).

For the efficient characterization of plant-associated viromes, there is generally a need to enrich viral sequences and conversely reduce the amount of host plant sequences that are generated. Different target nucleic acid populations have been used for virome studies but, coupled with the virus enrichment constraint, the most widely used approaches have targeted virion-associated nucleic acids (VANA) or double stranded RNAs (dsRNAs) (Maclot *et al.*, 2020). A huge number of bioinformatic tools are available for HTS data analysis and have been, together with nucleic acid preparation strategies, extensively reviewed (Roossinck *et al.*, 2015; Villamor *et al.*, 2019; Kutnjak *et al.*, 2021). The choice of a specific viral enrichment method or bioinformatic pipeline depends on the experimental objectives. For single plant samples or low complexity samples, the use of total RNA or siRNA sequencing might be the most universal and straightforward (Kreuze *et al.*, 2009; Kashif *et al.*, 2012) but when the viromes of entire plant communities are analyzed from complex plant pools, VANA or dsRNAs enrichment approaches are generally preferred (Thapa *et al.*, 2015; Bernardo *et al.*, 2018; Maclot *et al.*, 2020; Ma *et al.*, 2021). Even though there have been some efforts towards performance

comparisons of different virome analysis approaches, there is a need to better benchmark them and assess their respective efficiency at faithfully and comprehensively describing the complex viromes analyzed without introducing unwanted biases. In a virus discovery study on single quarantine plants, VANA was shown to assemble longer contigs compared to siRNA for a novel DNA virus (Mastrevirus) (Candresse *et al.*, 2014), while in a study investigating the virome of native plants in Oklahoma, more viral Operational Taxonomical Units (OTUs) could be detected with dsRNA compared to the VANA approach (Thapa *et al.*, 2015). Ma *et al.* (2019) provided a more comprehensive comparison of these two approaches using the natural viral communities present in complex plant pools from managed and unmanaged sites. The authors found significant differences with more viral contigs and, on average, longer contigs assembled from libraries prepared using the dsRNA approach. With regard to viral richness, more OTUs were detected by the dsRNA approach compared to the VANA approach. However, most DNA viruses were only detected using VANA. Standardization is fundamental for the reliable representation of microbiome/virome in metagenomic/phytovirome studies and is challenged by the rapid development of sequencing platforms, protocols and bioinformatic pipelines. Benchmarking is a powerful tool to provide standards that can be used to compare and evaluate the performance of the different steps required in metagenomic studies, including target nucleic acids population extraction, library preparation, sequencing (and sequencing platform) and finally bioinformatics sequence analysis.

In this context, benchmarking studies in metagenomics are often based on mock communities that are microbial assemblages of known composition which can be used to compare the actual *vs* the expected performance of a process. Besides the use of actual empirical phytoviromes (Ma *et al.* 2019), the use of synthetic communities could therefore provide a more precise and detailed benchmarking of HTS-based virome description strategies. Bacterial and fungal mock communities have thus been developed and used to compare the performance of different

sequencing platforms, e.g. short read Illumina or long read PacBio SMRT sequencing (Kozich *et al.*, 2013; Egan *et al.*, 2018, Sevim *et al.*, 2019). In recent years, viral mock communities have also been developed, especially in the medical and clinical field, to benchmark protocols in human virome studies. For example, the nucleic acid preparation step for the virome analysis of fecal samples was optimized using a combination of both viral and bacterial mock communities (Conceição-Neto *et al.*, 2015). In another study, the bias introduced by viral enrichment or random amplification were assessed using a DNA virus mock community (Parras-Moltó *et al.*, 2018). Viral synthetic communities have also been used to benchmark library preparation approaches in environmental (Roux *et al.*, 2016b) and insect (Gil *et al.*, 2021) virome studies. However, the use of synthetic communities in plant virome studies is lagging behind. The only study so far using a defined mix of plant viruses to assess different nucleic acid preparation protocols was performed by Gafaar and Ziebell (2020). This study revealed a better performance of enriched dsRNAs as compared to ribo-depleted total RNA or siRNAs for virus detection. However, so far only low complexity synthetic communities have been used, whereas most of the natural viral metagenomes are composed of a complex and diverse mixture of DNA and RNA viruses. In the present work, we used synthetic phytoviruses communities of varying degrees of complexity to compare the performance of VANA and dsRNA-based approaches for virome description and analyse how this performance is affected by sequencing depth and other parameters.

MATERIALS AND METHODS

Mock viral communities design

A list of 61 different viruses (60 different genera and 18 different families) was selected among those kept into collection and available at the German Collection of Microorganisms and Cell Cultures (Leibniz-Institute DSMZ, Braunschweig, Germany), taking into consideration a range of criteria. These criteria included (1) maximizing viral diversity by including viruses with all

genome types (ssDNA, dsDNA-RT, dsRNA, +ssRNA, -ssRNA) and including only a single representative virus per viral genus and (2) availability of a complete or near complete genomic sequence for each viral isolate included. In some cases, these genomic sequences had been determined previously while in other cases they were developed specifically in the frame of efforts to further improve the characterization of isolates distributed by the DSMZ through the EU-funded EVA-Global initiative (https://www.european-virus-archive.com/). Quality controlled samples were obtained from the DSMZ in the form of infected, lyophilized plant material in vacuum-sealed vials. The complete list of the isolates used, together with their properties and the propagation host in which they were provided are given in Table 1.

To create virus mock communities with different degrees of community complexity, initial low complexity pools were generated by assembling 30 mg of virus infected sample into 12 viral communities comprising five viruses each (150 mg of plant material each). Each of these communities contained at least one virus with a genome type different from +ssRNA (Supplementary Table S1).

Double-stranded RNA extraction

Double-stranded RNAs were purified from pooled samples according to the protocol of Marais *et al.* (2018) with some minor modifications. Briefly, instead of 75 mg, 150 mg dried plant material (representing a pool of five plants, Table S1) was used as starting material and buffer volumes increased proportionally. Plants were ground in liquid nitrogen until a fine powder was obtained which was then mixed with the phenol-extraction buffer. Following gentle agitation for 30 min and centrifugation, the supernatant was decanted and half of it directly further processed, while the other half was used for the stepwise gradual assembly of pairs of communities used to generate more complex viral communities. In this way, six communities of 10 viruses each, then three communities of 20 viruses and finally a single community of 60 viruses could be assembled. Between each step, assembled samples were vortexed for at least

30s for optimal homogenization. A detailed scheme of the pooling strategy to form communities of different complexities is shown in Supplementary Figure S1. Irrespective of its complexity, a supernatant volume corresponding to an initial input of 75 mg of plant samples was thus obtained and further processed as per the protocol of Marais *et al.* (2018) which involves two rounds of CC41 cellulose (Whatman) chromatography followed by a nuclease treatment (DNase RQ1 plus RNaseA under high salt conditions) to remove any remaining host DNA and single-stranded RNA. A negative extraction control using only buffer was systematically included. Purified dsRNAs were finally converted to cDNA and randomly amplified while simultaneously adding MID tags (François *et al.*, 2018a; Marais *et al.*, 2018).

VANA extraction

VANA extractions were performed on pools of five viruses similarly prepared as for dsRNA using the protocol of François *et al.* (2018a) with minor modifications. Briefly, 150 mg of lyophilized plant material (representing a pool of five plants, Table S1) were ground in Hank's buffered salt solution (HBSS) (1:10) with four metal beads. Following centrifugation, the supernatants were split and used in the same stepwise assembly of more complex communities as for the dsRNA approach (Supplementary Figure S1). A negative, buffer only, extraction control was systematically included. Each of the thus generated samples, representing different degrees of community complexity, was filtered through a 0.45-µm filter and centrifuged at 150,000g for 2.5 hours at 4°C to concentrate the virus particles. Unprotected nucleic acids were eliminated by DNase and RNase treatment at 37°C for 1.5 hours. Viral RNA and DNA were then isolated using the NucleoSpin Virus kit (Macherey Nagel, Hoerdt, France), using only 80 µl of sample in the first lysis step and omitting the addition of proteinase K. Extracted RNAs were transformed to cDNA using Superscript III reverse transcriptase (ThermoFisher Scientific/Invitrogen), cDNAs purified with the QIAquick PCR purification Kit (QIAGEN, Courtaboeuf, France) and a complementary strand was synthesized using the Klenow fragment

of DNA polymerase I before a random PCR amplification adding barcoded dodeca-linkers and corresponding MID primers during reverse transcription and PCR, respectively (François *et al.*, 2018a)

Illumina sequencing

PCR products from all communities analyzed using the dsRNA and VANA procedures were finally purified using the MinElute PCR purification kit (QIAGEN) and equimolar quantities of amplification products were sent to Illumina sequencing in multiplexed format (2 × 150 bp) on two lanes (one for VANA and one for dsRNA, respectively) on a NovaSeq 6000 system at the GetPlaGe platform (GenoToul INRAE Toulouse, France).

HTS data analysis

Sequencing reads were imported into CLC Genomics Workbench v. 21.0.3. (CLC-GW) and adapters were removed from reads followed by trimming on quality and length using default settings and a minimum read length of 60 nucleotides (nt). Final trimmed reads were on average 111-113 nt long for the various datasets. Datasets were normalized by resampling at varying depth as needed, using the random reads sampling tool in CLC-GW.

To analyse virus detection performance as a function of contig size, *de novo* assembly was performed with CLC-GW (word size, 50; bubble size, 300) using various minimum contig lengths (125, 175, 250, 350, 500, 1000 nt). In order to identify viruses that were not expected in the analyzed samples, contigs were annotated by a BlastX analysis (Altschul *et al.*, 1990) against the viral RefSeq portion of the non-redundant (nr/nt) NCBI GenBank database. For unexpected viruses thus identified, a genomic scaffold was reconstructed and extended by repeated rounds of residual reads mapping using CLC-GW, thus yielding near complete genome sequences that were used as reference for the relevant virus (Table 2). In a few cases, these assemblies were considered too incomplete and the closest complete genomic sequence in GenBank was selected as reference sequence (Table 2).

In order to determine virus detection performance, unassembled reads or *de novo* assembled contigs were mapped against the reference genome segment(s) for each virus (Tables 1 and 2) using very stringent mapping parameters (length fraction 100%, minimal similarity fraction 90%) in CLC-GW. In order to take into account inter-sample crosstalk due to index jumping (Illumina, 2017; van der Valk *et al.*, 2019), a threshold of positive detection was computed for each viral molecule by calculating the average + 3 standard deviations (SD) of background reads in libraries generated from communities that did not contain the corresponding virus.

RESULTS

Expected and unexpected viruses or virus-like agents in viral communities HTS data

The analysis of reads from both the VANA and dsRNA approaches for all communities revealed the presence of all expected viruses, although a few viruses were only represented by a limited number of reads or were only detected using one of the two approaches. Overall, only lettuce ring necrosis virus turned out to be fully absent from VANA reads, while banana bunchy top virus was only represented by a single dsRNA read. It should also be noted that not all viruses could be detected in all the communities of different complexity in which they were present.

In addition to the expected 61 viruses, evidence for the presence in some communities of unexpected viruses or virus-like agents was obtained through the BlastX indexing of *de novo* assembled contigs from the low complexity, 5-viruses communities. A total of 11 unexpected agents were thus identified (Table 2). These include three linear ssRNA satellites associated with the helper virus isolates included in the communities [turnip crinkle satellite F (TCVsatRNA F), pea enation mosaic satellite RNA (PEMVsatRNA) and strawberry latent ringspot virus satellite RNA (SLRSVsatRNA)], latent viruses associated with the propagation hosts used [Hordeum vulgare endornavirus (HvEV), maize-associated totivirus (MATV), maize-associated totivirus2 (MTV-2) and Chenopodium quinoa mitovirus 1 (CqMV1)], as well

as viruses in coinfection with some viral isolates used [poinsettia mosaic virus (PnMV), tobacco mosaic virus (TMV), turnip yellows virus (TuYV) and maize streak Réunion virus (MSRV)] (Table 2). For these additional unexpected viruses, either a nearly complete genome was reconstructed from sequencing reads and used as the mapping reference or the closest full genome sequence in GenBank was used for further mapping analyses (Table 2). For all other viral isolates included in the communities, complete or nearly complete genomic sequences are available (Table 1).

While the communities of varying complexities analysed here will be referred to a 5-viruses, 10-viruses, 20-viruses and 60-viruses, it should be kept in mind that the real number of the viruses present in a given community might be slightly different because of (i) the presence of one or more of the unexpected viruses and (ii) the counting of pea enation mosaic virus as one virus when it is in fact a co-infection of pea enation mosaic virus 1 (*Enamovirus*) and pea enation mosaic virus 2 (*Umbravirus*).

Read mapping analysis of VANA and dsRNA datasets for the communities of various complexities

To be able to compare results between low and high complexity communities, all datasets were normalised by randomly subsampling 120K cleaned reads. To adress the issue of inter-sample crosstalk caused by index jumping (Illumina, 2017; van der Valk *et al.*, 2019) a threshold of positive detection was computed for each viral molecule by calculating the average + 3 standard deviations (SD) of background reads in libraries generated from communities that did not contain the corresponding virus. Assuming a normal distribution of crosstalk reads numbers, this strategy ensures that the probability of having a mapped reads number higher than the threshold by chance (false positive detection) has a probability of less than 1%.

In general, the proportion of viral reads was high (64-89%) in both VANA and dsRNA datasets and was slightly affected by community complexity, with a general trend to reach a higher

proportion of viral reads when analysing more complex communities (Figure 1A). The proportion of viral reads in the dsRNA datasets were slightly higher than in the corresponding VANA datasets, with the strongest effect observed for lower complexity, 5- and 10-viruses communities (64-65% viral reads as compared to 79-82%, Figure 1A).

Using the 12 communities of five viruses and a sequencing depth of 120K reads, 67 viruses were detected with both VANA and dsRNA approaches (with detection of reads for a single genomic molecule considered as positive detection for a virus with a divided genome), out of the total 72 viruses or virus-like agents in the 12 communities analysed (93.1%). However, VANA yielded reads for all six DNA viruses present, when dsRNA yielded reads for only three of them. Conversely, VANA yielded reads for 61 of the 66 RNA viruses or satellites (92.4%), when dsRNA yielded reads for 64 of them (97.0%) (Figure 1B). As expected and previously reported, the performance of VANA is thus superior for DNA viruses but that of dsRNA superior for RNA viruses.

The impact of increasing community complexity is reflected by the diminishing number of viruses detected at an equal sequencing depth effort of 120K reads. The performance of VANA gradually deteriorated, going from 61 viruses detected to 58 (10-viruses communities) and then to 52 (20-viruses communities) to reach only 34 viruses detected (51.5%) with the most complex community (Figure 1B). The same pattern is observed for DNA viruses, where all six DNA viruses were detected for the 10- and 20-viruses communities but only one DNA virus detected when analysing the 60-viruses community. In the case of the dsRNA approach, performance was marginally reduced for 10- and 20-viruses communities (65 and 63 viruses detected, respectively) and less affected than for the VANA approach for the most complex community with still 57 of 66 viruses detected (86.4%) (Figure 1B).

However, if trying to compensate for community complexity by proportionally increasing the sequencing effort for more complex communities, the erosion in performance is strongly

reduced for VANA, with still 57 of 66 RNA viruses detected for the 60-virus community (86.4%) sequenced at 1.44 M reads depth (12 x 120K) and five of the six DNA viruses (83.3%). The performance of dsRNA, on the other hand, is no longer impaired, as all 66 RNA viruses (100%) are still detected for the most complex community (result not shown).

The stronger degradation of VANA performance as community complexity increases, correlates with a more uneven distribution of read numbers between viruses and the stronger dominance of a few viruses, in particular turnip yellow mosaic virus (TYMV). In the 60-viruses community VANA dataset, TYMV represented 67% of the reads while the corresponding value for the dsRNA dataset was only 28%. As shown in Figure 2, even if spanning a 5 to 6 logs scale, read numbers (or proportion of reads over total dataset reads) tend to be somewhat more evenly distributed between viruses in the dsRNA dataset than in the VANA dataset for the 60-viruses community.

Although allowing to compare the performance of the VANA and dsRNA approaches, these analyses based on the mapping of reads against cognate reference genomes do not mimic the situation in metagenomic studies, in which a high proportion of viruses are expected to be novel and for which therefore no suitable reference genome is available. We therefore analysed the performance of these two approaches following a *de novo* assembly of reads into contigs, which is known to reduce the proportion of un-annotated “dark matter” (François *et al.*, 2018b).

Impact of minimal contig length on the number of detected viruses

We first evaluated the impact of the minimal contig length on the number of detected viruses using the most complex community of 60 viruses and deep datasets normalized at 10 million reads. As expected and as seen in Figure 3, the number of viruses reported decreased as minimal contig length increased. The pattern observed in the figure for RNA viruses is also observed for DNA viruses (not shown). The dsRNA approach consistently detected more viruses than the VANA one, irrespective of the minimal contig length used, but the difference increased as

minimal contig length increased. Using the shortest, 125 nt contig length, VANA identified 54 of the 66 RNA viruses present in the community (81.8%), while dsRNA identified 63 of them (95.5%) (Figure 3). The corresponding values for DNA viruses are respectively 4 (66.7%) and 3 (50%).

On the other hand, the coverage of the detected viruses (fraction of the target molecules represented in contigs) was much less affected by minimal contig length. Relatively stable for the dsRNA approach, for which it varied between 66.5% and 74.9% with no clear trend, it showed a tendency to increase with contig length for the VANA approach, from 50.2% (>125 nt contigs) to 76.7% (>1,000 nt contigs) (Supplementary Figure S2).

For further analyses, an intermediate 250 nt minimal contig length was retained as it corresponds to an encoded 83 amino acids sequence that was felt sufficient for many conserved protein domain searches which are often used in virome analysis or annotation (Lefebvre *et al.*, 2019).

Effects of community complexity on virome description performance

We evaluated how, for a given sequencing depth, community complexity affects virome description performance following contigs assembly. For this, all datasets were normalized at a 120k read depth. Similar to the situation observed using reads mapping (Figure 2), the number of detected viruses was reduced as community complexity increased. Again, dsRNA outperformed VANA at all complexity levels and VANA performance degradation was more drastic at high community complexity, dropping from 44 viruses detected for communities of five viruses (66.7% of total viruses) to 11 viruses detected (16.7%) for the 60-viruses community. The corresponding values for dsRNA are 53 viruses (80.3%) and 26 viruses (39.4%) (Supplementary Figure S3). These results indicate that even for limited complexity communities involving only five viruses, reads numbers significantly higher than 120K are

needed by both techniques to achieve a 100% detection performance with a wide range of viruses.

If trying to compensate increased virome complexity by a parallel increase in sequencing depth, a negative impact of complexity is still seen but is much less severe. For example, for the most complex community of 60 viruses, at a 1.44M depth (12*120K reads), VANA detected 23 RNA viruses (compared to 44 when analysing individually the 12 pools of five viruses at 120K reads depth) a reduction in performance of 47.8%, while dsRNA detected 42 viruses to be compared with 53 viruses when individually analysing the 12 pools of five viruses, a reduction in performance of 20.7% (Figure S4). The loss in performance resulting from community complexity is therefore significant for both approaches, but stronger for VANA.

Impact of sequencing depth on *de novo* assembly

The 60-viruses community was used to investigate the influence of sequencing depth on *de novo* assembly performance itself. The VANA and dsRNA datasets were therefore resampled at different depths (100K, 300K, 1M, 3M and 10M reads, five random resampling at each depth), assembled and mapped against the viral reference genomes to determine the average assembly parameters and viral contigs parameters. The results are shown in Supplementary Table S2 and, for viral contigs alone, in Table 3.

As expected, all assembly parameters (number of contigs, average contig length, N50, maximal contig length) increased with sequencing depth (Supplementary Table S2). The same tends to be true for viral contigs (number and length, Table 3), while the proportion of viral contigs tended to diminish as sequencing depth increased, likely reflecting increased probability of assembly of non-viral reads (Supplementary Table S2). Although at the lowest, 100K reads sequencing depth, few assembly parameters were found to be statistically different, both the total number of assembled contigs and the number of viral contigs were found to be highly

statistically different, with dsRNA yielding about 3-fold more contigs and 3-fold more viral contigs than VANA (Table 3 and Table S2). This trend was observed at all sequencing depth, with 1.3 to 1.8-fold more viral contigs identified for dsRNA.

At other sequencing depths, differences between the VANA and dsRNA assemblies proved systematically highly significant, with dsRNA consistently yielding more numerous and longer contigs as well as more numerous and longer viral contigs. On the other hand, the proportion of viral contigs was found consistently higher in the assemblies of the VANA datasets (Supplementary Table S2).

It should be noted that the better performance of dsRNA is independent of the minimal contig length (Table 4). In particular, using the most complex community and 10 million reads datasets, the higher performance of dsRNA over VANA was observed for all assembly parameters (number of contigs, average length, N50, maximum length) and for all viral contigs parameters (number and average length) at all minimal contigs length (from 125 to 1000 nt) with a single exception, the number of viral contigs >125 nt long (1,852 for VANA vs 1,672 for dsRNA) (Table 4). At all other minimal contig length, VANA showed from 19.2% (contigs ≥175 nt) to 50.7% (>1 kb contigs) fewer viral contigs than dsRNA and these contigs were 23-33% shorter on average than the dsRNA ones (Table 4).

Impact of sequencing depth on virus identification performance

We proceeded to evaluate the performance of VANA and dsRNA in identifying the expected viruses or viral molecules as affected by sequencing depth. The contigs obtained for the various datasets resampled at different depths (five resamplings per sequencing depth) were mapped on individual reference sequences. This allowed to evaluate both the proportion of detected viruses and the coverage of the detected viral molecules, together with their standard deviation (Supplementary Figure S5). Once again, at all sequencing depths and for both parameters,

dsRNA outperformed VANA for RNA viruses, while VANA outperformed dsRNA for DNA viruses. In all cases, average coverage of detected molecules showed a high standard deviation but dsRNA contigs covered 9% to 22% more of the detected molecules than VANA contigs.

Similarly, and as expected from single reads mapping data, dsRNA outperformed VANA for the identification of RNA viruses present in the most complex, 60-viruses community. For VANA, performance ranged from 17.7% of RNA viruses identified at the 100K reads depth to 60.3% at the 10 million reads depth. The corresponding values for dsRNA are respectively 35.2% and 89.7%.

A plot of the observed proportion of detected viruses over a logarithmic scale of the sequencing effort is shown in Figure 4. It shows a remarkable pattern with linear regression r^2 coefficients of 0.97-0.99 suggesting a very strong and monotonous relationship between sequencing depth and the proportion of the viruses present in the community that are represented by at least one assembled contig. An extension of that trend would suggest that a depth of about 30 million reads would be needed for the dsRNA approach to recover at least one contig for each of the 66 RNA viruses present in the synthetic community, while in excess of 1 billion reads would be need to achieve a comparable performance using VANA.

Due to a more limited number of reads available for virus communities up to the 20-viruses pools, a similar evaluation could not be as extensively performed for these lower complexity communities. However, an analysis at three sequencing depths (100K reads, 300K reads, 875K reads) of the 20-viruses communities data provided comparable results with r^2 correlation coefficients of 0.95-0.98, suggesting that the linear correlation between the percentage of viruses recovered and the log of the sequencing depth is independent of the complexity of the analysed community (result not shown).

An analysis performed at the level of individual viral genomic molecules (115 viral molecules) allows to evaluate the performance of the two methods using the most complex, 60-viruses

pool, for groups of viruses with different genome types. The numbers are however small for RNA satellites, dsRNA viruses and dsDNA viruses. The results, using a 10 million reads sequencing depth, are summarized in Table 5. Considering individual molecules, VANA had at least one contig for only 50% of the viral molecules present in the most complex synthetic community, to be compared with a 76.5% value for dsRNA. But while the VANA performance was at an intermediate level for all virus groups analysed, dsRNA showed good performance for +ssRNA viruses (89.5% of molecules), RNA satellites (100%) and dsRNA viruses (100%). As expected, dsRNA performance was however poor for DNA viruses but also for -ssRNA viruses (41.7% of detected molecules only).

DISCUSSION

While synthetic communities have been widely used to benchmark metagenomic processes targeting bacteria and fungi, methodological benchmarking approaches in virome studies are still more limited. Although progress has been made in standardizing viral metagenomic approaches in clinical settings (Ajami *et al.*, 2018, Parras-Moltó *et al.*, 2018; Santiago-Rodriguez *et al.*, 2020) and, to some extent, in environmental virome studies (Roux *et al.*, 2016a; Zablocki *et al.*, 2021), such approaches are largely lacking in plant virology. Here we used DSMZ quality-controlled plant samples that allowed for a simplified construction of synthetic viral communities of varying complexity. Although some of the viruses were detected by only very low read numbers, no virus was fully absent from all generated datasets, validating the approach and the samples used. The fact that some viruses were identified only by low read numbers could have a variety of reasons, such as low virus titer in some samples, competition with other viruses for representation in the assembled communities, or difficulties in extracting viruses from certain host plant species. In this respect, it should be noted that the two viruses present as infected banana samples, banana streak OL virus (BSOLV) and banana bunchy top virus (BBTV) were only detected by very low read numbers in both techniques used, despite

the fact that these techniques have successfully been used in the past to analyse banana samples (Teycheney *et al.*, 2015; Filloux *et al.*, 2015).

On the other hand, a few unexpected viruses or viral agents were identified. In most cases these correspond to satellites that had not been specifically indexed in the isolates used or of viruses latently infecting propagation hosts, such as Hordeum vulgare endornavirus, which is present in many barley varieties or Chenopodium quinoa mitovirus.

The communities assembled cover all known phytovirus genome types, 21 viral families (plus satellites and one virus unassigned in a family) and a total of 62 genera [plus two viruses not currently assigned to a genus (poinsettia mosaic virus and spartina mottle virus) and three satellites]. It is thus probably to date the largest scale effort to build synthetic viral communities and use them for the benchmarking of phytovirome analysis approaches. In some benchmarking studies, the nucleic acid proportions of the individual viruses involved in the virus community were quantified prior to extraction (Roux *et al.*, 2016b, Sevim *et al.*, 2019). The fact that, in our study, no special effort was made to normalize or measure the concentration of the different viruses is a limitation for some comparisons. On the other hand, the samples used involved different propagation hosts and actual virus titres in those hosts, so that the communities assembled are more likely to reflect actual samples from phytovirome studies. The results obtained indicate that a range of parameters impact the completeness of the virome description achieved. Not surprisingly, such parameters include (i) sequencing depth, (ii) community complexity, (iii) use of *de novo* assembled contigs *vs* use of unassembled reads and (iv) minimal contig length.

As was previously reported using natural communities (Ma *et al.*, 2019), the dsRNA approach provided in all comparisons a more complete description of the RNA virome than the VANA approach but performed very poorly with DNA viruses. However, the differential with VANA is more limited for the less complex, 5- or 10-viruses communities. The dsRNA approach is

therefore recommended whenever analysing complex viromes or when an emphasis on RNA viruses is of importance, in particular since it allows comparable levels of completeness with a lower sequencing effort. The reason for the better performance of the dsRNA approach is not fully clear but might result from a lower level of competition between viruses for representation in complex pools, resulting in the somewhat less imbalanced distribution of read numbers between viruses (Figure 2). Different human microbiome studies have shown that different steps of RNA/DNA extraction such as homogenization, centrifugation, filtration and chloroform treatment, can have a major impact on the quantitative and qualitative composition of identified virus communities, skewing viral assemblages in metagenomes (Kleiner *et al.*, 2015; Conceição-Neto *et al.*, 2015; Parras-Moltó *et al.*, 2019). Another critical step is the library preparation step, that often involves a random amplification PCR to increase virus genetic material and to add linkers allowing the multiplexing samples during HTS sequencing, thus reducing sequencing costs. The amplification step alters the relative abundance of viruses and can lead to uneven coverage if random primers do not anneal randomly on viral genomes. Indeed, in the case of faba bean necrotic stunt virus, the relative frequencies of the different genome segments determined by qPCR was significantly different before and after a rolling circle amplification step used prior to HTS sequencing (Gallet *et al.*, 2017). Furthermore, different library preparation techniques have been found to require different sequencing depths to achieve the same genome coverage (Visser *et al.*, 2016). Regardless of the experiment, it is advisable to develop an estimate of the sequencing depth needed, so as to be able to answer the biological question at hand while avoiding excessive sequencing costs. Here we identified a very robust correlation between the percentage of viruses identified in complex communities and the log of the sequencing depth. This is an interesting result, since it allows to gauge the sequencing effort needed for a particular level of virome description or, conversely, to gauge the extent of virome description that can be expected from a particular sequencing depth.

Besides metagenomic studies, this finding might have practical implications for diagnostics since many plants, in particular vegetatively propagated ones, frequently display complex mixed infections involving a range of viruses.

Virus detection in metagenomic studies is constrained by the degree of complexity of the virus communities analyzed. Our results suggest that the detection efficiency of either mapping of unassembled read or analysis of *de novo* assembled contigs were affected by community complexity with a general trend of detecting a lower proportion of viruses in more complex communities. However, the read mapping strategy was more efficient at all complexities (Figure 1B and Supplementary Figure S3), confirming results obtained through performance testing of sequence analysis strategies (Massart *et al.*, 2019). This may be due to the complexity of *de novo* assembly of complex communities, linked with insufficient coverage or uneven coverage of low abundance viruses within such communities. Correspondingly, we observed a lower virus detection rate when using longer minimal contig sizes in the *de novo* assembly, which again might be attributed to difficulties in assembling reads from more complex communities for example when coexisting viruses share highly similar regions in their genomes, leading to higher fragmentation and reduced contig sizes (Roux *et al.*, 2017).

Lastly, it has been reported that the quality and completeness of virome description is also affected by the analysis pipeline used (Rampelli *et al.*, 2016; Roux *et al.*, 2017; Breitwieser *et al.*, 2019, Sutton *et al.*, 2019). The normalized 10M reads datasets generated in the present study with the 60-viruses community, which have been made available, together with the community composition and the complete or near complete genomic sequences of the isolates involved should prove very useful tools to benchmark virome characterization pipelines.

Acknowledgements: The authors wish to thank the INRAE GetPlaGe Platform (GenoToul, Toulouse, France) for Illumina sequencing.

Funding: This study was funded by the European Union through a Horizon 2020 Marie Skłodowska-Curie Actions Innovative Training Network (H2020 MSCA- 60 ITN) project "INEXTVIR" (GA 813542).

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

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

REFERENCES

Ajami NJ, Wong MC, Ross MC, Lloyd RE, Petrosino F (2018). Maximal viral information recovery from sequence data using VirMAP. Nat. Comm. 9, 3205. https://doi.org/10.1038/s41467-018-05658-8

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990) Basic local alignment search tool. J Mol Biol. 215 403-410. https://doi: 10.1016/S0022-2836(05)80360-2

Anderson PK, Cunningham AA, Patel NG, Morales FJ, Epstein PR, Daszak P (2004) Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. Trends Ecol. Evol. 19, 535-544. https://doi.org/10.1016/j.tree.2004.07.021

Bernardo P, Charles-Dominique T, Barakat M, Ortet P, Fernandez E, Filloux D, Hartnady P, Rebelo T A, Cousins SR, Mesleard F, Cohez D, Yavercovski N, Varsani A, Harkins GW, Peterschmitt M, Malmstrom CM, Martin DP, Roumagnac P (2018) Geometagenomics illuminates the impact of agriculture on the distribution and prevalence of plant viruses at the ecosystem scale. ISME J. 12, 173-184. https://doi.org/10.1038/ismej.2017.155

Breitwieser FP, Lu J, Salzberg SL (2019) A review of methods and databases for metagenomic classification and assembly. Brief Bioinform. 20, 1125-1136. https://doi.org/10.1093/bib/bbx120

Candresse T, Filloux D, Muhire B, Julian C, Galzi S, Fort G, Bernardo P, Daugrois JH, Fernandez E, Martin DP, 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. https://doi.org/10.1371/journal.pone.0102945

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

Conceição-Neto N, Zeller M, Lefrère H, De Bruyn P, Beller L, Deboutte W, Yinda CK, Lavigne R, Maes P, Ranst M Van, Heylen E, Matthijnssens J (2015) Modular approach to customise sample preparation procedures for viral metagenomics: A reproducible protocol for virome analysis. Sci. Rep. 5, 1-14. https://doi.org/10.1038/srep16532

Egan CP, Rummel A, Kokkoris V, Klironomos J, Lekberg Y, Hart M (2018) Using mock communities of arbuscular mycorrhizal fungi to evaluate fidelity associated with Illumina sequencing. Fungal Ecol. 33, 52-64. https://doi.org/10.1016/j.funeco.2018.01.004

Filloux D, Dallot S, Delaunay A, Galzi S, Jacquot E, Roumagnac P (2015) Metagenomics approaches based on virion-associated nucleic acids (VANA): An innovative tool for assessing without a priori viral diversity of plants. Meth. Molec. Biol. 1302, 249-257. https://doi.org/10.1007/978-1-4939-2620-6_18

François S, Filloux D, Fernandez E, Ogliastro M, Roumagnac P (2018a) Viral metagenomics approaches for high-resolution screening of multiplexed arthropod and plant viral communities. In: Viral metagenomics: methods and protocols, Pantaleo V, Chiumenti M (Eds), Humana Press, New York (USA), pp 77-95.

François S, Filloux D, Frayssinet M, Roumagnac P, Martin DP, Ogliastro M, Froissart R (2018b) Increase in taxonomic assignment efficiency of viral reads in metagenomic studies. Virus Res. 244, 230-234. https://doi.org/10.1016/j.virusres.2017.11.011

Gaafar YZA, 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, 1-17. https://doi.org/10.1371/journal.pone.0237951

Gallet R, Fabre F, Michalakis Y, Blanc S (2017) The number of target molecules of the amplification step limits accuracy and sensitivity in ultradeep-sequencing viral population studies. J Virol. 91, e00561-17. https://doi.org/10.1128/JVI.00561-17

Gil P, Dupuy V, Koual R, Exbrayat A, Loire E, Fall AG, Gimonneau G, Biteye B, Talla Seck M, Rakotoarivony I, Marie A, Frances B, Lambert G, Reveillaud J, Balenghien T, Garros C, Albina E, Eloit M, Gutierrez S (2021) A library preparation optimized for metagenomics of RNA viruses. Mol. Ecol. Res. 21, 1788-1807. https://doi.org/10.1111/1755-0998.13378

Hasiów-Jaroszewska B, Boezen D, Zwart MP (2021) Metagenomic studies of viruses in weeds and wild plants: a powerful approach to characterise variable virus communities. Viruses 13, 1939. https://doi.org/10.3390/v13101939.

Illumina (2017) Effects of index misassignment on multiplexing and downstream analysis. https://www.illumina.com/content/dam/illumina-marketing/documents/products/whitepapers/index-hopping-white-paper-770-2017-004.pdf

Jian H, Yi Y, Wang J, Hao Y, Zhang M, Wang S, Meng C, Zhang Y, Jing H, Wang Y, Xiao X (2021) Diversity and distribution of viruses inhabiting the deepest ocean on Earth. ISME J. 15, 3094-3110. https://doi.org/10.1038/s41396-021-00994-y

Kashif M, Pietila S, Artola K, Jones RAC, Tugume AK, Mäkinen V, Valkonen JPT (2012) Detection of viruses in sweet potato from Honduras and Guatemala augmented by deep-sequencing of small RNAs. Plant Dis. 96, 1430-1437. https://doi.org/10.1094/PDIS-03-12-0268-RE

Kleiner M, Hooper LV, Duerkop BA (2015) Evaluation of methods to purify virus-like particles for metagenomic sequencing of intestinal viromes. BMC Genomics 16, 7. https://doi.org/10.1186/s12864-014-1207-4

Kobayashi K, Atsumi G, Iwadate Y, Tomita R, Chiba K, Akasaka S, Nishihara M, Takahashi H, Yamaoka N, Nishiguchi M, Sekine K (2013) Gentian Kobu-sho-associated virus: a tentative, novel double-stranded RNA virus that is relevant to gentian Kobu-sho syndrome. J. Gen. Plant Pathol. 79, 56-63. https://doi.org/10.1007/s10327-012-0423-5

Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq illumina sequencing platform. Appl. Environ. Microbiol. 79, 5112-5120. https://doi.org/10.1128/AEM.01043-13

Kreuze JF, 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. https://doi.org/10.1016/j.virol.2009.03.024

Kutnjak D, Tamisier L, Adams I, Boonham N, Candresse T, Chiumenti M, De Jonghe K, Kreuze JF, Lefebvre M, Silva G, Malapi-Wight M, Margaria P, Mavrič Pleško I, McGreig S, Miozzi L, Remenant B, Reynard JS, 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. https://doi.org/10.3390/microorganisms9040841

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

Lefeuvre P, Martin DP, Elena SF, Shepherd DN, Roumagnac P, Varsani A (2019) Evolution and ecology of plant viruses. Nat. Rev. Microbiol. 17, 632-644. https://doi.org/10.1038/s41579-019-0232-3

Ma Y, Fort T, Marais A, Lefebvre M, Theil S, Vacher C, Candresse T (2021) Leaf-associated fungal and viral communities of wild plant populations differ between cultivated and natural ecosystems. Plant Environ. Interact. 2, 87-99. https://doi.org/10.1002/pei3.10043

Ma Y, Marais A, Lefebvre M, Faure C, Candresse T (2020) Metagenomic analysis of virome cross-talk between cultivated Solanum lycopersicum and wild Solanum nigrum. Virology 540, 38-44. https://doi.org/10.1016/j.virol.2019.11.009

Ma Y, Marais A, Lefebvre M, Theil S, Svanella-Dumas L, Faure C, Candresse T (2019) Phytovirome analysis of wild plant populations: comparison of double-stranded RNA and virion-associated nucleic acids metagenomic approaches. J. Virol. 94, e01462-19. https://doi.org/10.1128/jvi.01462-19

Maachi A, Donaire L, Hernando Y, Aranda MA (2022) Genetic differentiation and migration fluxes of viruses from melon crops and crop edge weeds. J. Virol. 96, e00421-22. https://doi.org/10.1128/jvi.00421-22

Maclot F, Candresse T, Filloux D, Malmstrom CM, Roumagnac P, van der Vlugt R, Massart S (2020) Illuminating an ecological blackbox: using high throughput sequencing to characterize the plant virome across scales. Front. Microbiol. 11, 578064. https://doi.org/10.3389/fmicb.2020.578064

Malmstrom CM, Melcher U, Bosque-Pérez NA (2011) The expanding field of plant virus ecology: historical foundations, knowledge gaps, and research directions. Virus Res. 159, 84-94. https://doi.org/10.1016/j.virusres.2011.05.010

Marais A, Faure C, Bergey B, Candresse T (2018) Viral double-stranded RNAs (dsRNAs) from Plants: Alternative nucleic acid substrates for high-throughput sequencing. In: Viral Metagenomics: Methods in Molecular Biology, Pantaleo V, Chiumenti M (Eds.), Humana Press, New York (USA), pp 45-53.

Maree HJ, Fox A, Al Rwahnih M, Boonham N, Candresse T (2018) Application of HTS for routine plant virus diagnostics: state of the art and challenges. Front. Plant Sci. 9, 1082. https://doi.org/10.3389/fpls.2018.01082

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 HJ, Olivier T, Olmos A, Pooggin MM, Reynard JS, Ruiz-García AB, Safarova D, Schneeberger PHH, Sela N, Turco S, Vainio EJ, Varallyay E, Verdin E, Westenberg M, Brostaux Y, Candresse T

(2019) Virus detection by high-throughput sequencing of small RNAs: large-scale performance testing of sequence analysis strategies. Phytopathology 109, 488-497. https://doi.org/10.1094/PHYTO-02-18-0067-R

Moreno AB, López-Moya JJ (2020) When viruses play team sports: mixed infections in plants. Phytopathology 110, 29-48. https://doi.org/10.1094/PHYTO-07-19-0250-FI

Parras-Moltó M, Rodríguez-Galet A, Suárez-Rodríguez P, López-Bueno A (2018) Evaluation of bias induced by viral enrichment and random amplification protocols in metagenomic surveys of saliva DNA viruses. Microbiome 6,1-18. https://doi.org/10.1186/s40168-018-0507-3

Rampelli S, Soverini M, Turroni S, Quercia S, Biagi E, Brigidi P, Candela M (2016) ViromeScan: a new tool for metagenomic viral community profiling. BMC Genomics, 17, 1-9. https://doi.org/10.1186/s12864-016-2446-3

Roossinck MJ, Martin DP, Roumagnac P (2015) Plant virus metagenomics: advances in virus discovery. Phytopathology 105, 716-727. https://doi.org/10.1094/PHYTO-12-14-0356-RVW

Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, Poulos BT, Solonenko N, Lara E, Poulain J, Pesant S, Kandels-Lewis S, Dimier C, Picheral M, Searson S, Cruaud C, Alberti A, Duarte CM, Gasol JM, Vaqué D, Tara Oceans Coordinators, Bork P, Acinas SG, Wincker P, Sullivan MB (2016a) Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. Nature 537, 689-693. https://doi.org/10.1038/nature19366

Roux S, Emerson JB, Eloe-Fadrosh EA, Sullivan MB (2017) Benchmarking viromics: an *in silico* evaluation of metagenome-enabled estimates of viral community composition and diversity. Peer J. 5, e3817. https://doi.org/10.7717/peerj.3817

Roux S, Matthijnssens J, Dutilh BE (2021) Metagenomics in Virology. In Encyclopedia of Virology (Fourth Edition), Bamford DH and Zuckerman M (eds) Academic Press, Oxford, pp. 133–140. https://doi.org/10.1016/B978-0-12-809633-8.20957-6

Roux S, Solonenko NE, Dang VT, Poulos BT, Schwenck SM, Goldsmith DB, Coleman ML, Breitbart M, Sullivan MB (2016b) Towards quantitative viromics for both double-stranded and single-stranded DNA viruses. Peer J. 4, e2777. https://doi.org/10.7717/peerj.2777

Santiago-Rodriguez TM, Hollister EB (2020) Potential applications of human viral metagenomics and reference materials: considerations for current and future viruses. Appl. Environ. Microbiol. 86, e01794-20. https://doi.org/10.1128/AEM.01794-20

Schönegger D, Marais A, Faure C, Candresse T (2022) A new flavi-like virus identified in populations of wild carrots. Arch Virol. 167, 2407-2409. https://doi: 10.1007/s00705-022-05544-1

Sevim V, Lee J, Egan R, Clum A, Hundley H, Lee J, Everroad RC, Detweiler AM, Bebout BM, Pett-Ridge J, Göker M, Murray AE, Lindemann SR, Klenk HP, O'Malley R, Zane M, Cheng JF, Copeland A, Daum C, Woyke T (2019) Shotgun metagenome data of a defined mock community using Oxford Nanopore, PacBio and Illumina technologies. Sci. Data, 6, 285. https://doi.org/10.1038/s41597-019-0287-z

Shi M, Lin XD, Chen X, Tian JH, Chen LJ, Li K, Wang W, Eden JS, Shen JJ, Liu L, Holmes EC, Zhang YZ (2018) The evolutionary history of vertebrate RNA viruses. Nature 556, 197-202. https://doi.org/10.1038/s41586-018-0012-7

Susi H, Laine AL (2021) Agricultural land use disrupts biodiversity mediation of virus infections in wild plant populations. New Phytol. 230, 2447-2458. https://doi.org/10.1111/nph.17156

Sutton TDS, Clooney AG, Ryan FJ, Ross RP, Hill C (2019) Choice of assembly software has a critical impact on virome characterisation. Microbiome 7, 1-15. https://doi.org/10.1186/s40168-019-0626-5

Teycheney PY, Bandou E, Gomez RM, LangeD, Pavis C, Umber M, Acina Manbole IN, Bonheur L, Daugrois JH, Fernandez E, Filloux D, Julian C, Roumagnac P, Grisoni M, Pierret A, Rubington M, Candresse T, Contreras S, Faure C, Marais A, Theil S, Da Câmara MA, Mendonça D, Pinheiro de Carvalho M (2015). Viral treasure hunt in European outermost territories: how metagenomics boosts the discovery of novel viral species in tropical and sub-tropical crops germplasm. https://agritrop.cirad.fr/575812/

Thapa V, McGlinn DJ, Melcher U, Palmer MW, Roossinck MJ (2015) Determinants of taxonomic composition of plant viruses at the Nature Conservancy's Tallgrass Prairie Preserve, Oklahoma. Virus Evol. 1, 1-8. https://doi.org/10.1093/ve/vev007

van der Valk T, Vezzi F, Ormestad M, Dalén L, Guschanski K (2019) Index hopping on the Illumina HiseqX platform and its consequences for ancient DNA studies. Mol. Ecol. Resour. 20, 1171-1181 https://doi.org/10.1111/1755-0998.13009

Villamor DEV, Ho T, Al Rwahnih M, Martin RR, Tzanetakis IE (2019) High throughput sequencing for plant virus detection and discovery. Phytopathology 109, 716-725. https://doi.org/ 10.1094/PHYTO-07-18-0257-RVW

Visser M, Bester R, Burger JT, Maree HJ (2016) Next-generation sequencing for virus detection: covering all the bases. Virol J. 13, 85. https://doi.org/10.1186/s12985-016-0539-x

Wren JD, Roossinck MJ, Nelson RS, S2cheets K, Palmer MW, Melcher U (2006) Plant virus biodiversity and ecology. PLoS Biol. 4, e80. https://doi.org/10.1371/journal.pbio.0040080

Zablocki O, Michelsen M, Burris M, Solonenko N, Warwick-Dugdale J, Ghosh R, Pett-Ridge J, Sullivan MB, Temperton B (2021) VirION2: A shortand long-read sequencing and informatics workflow to study the genomic diversity of viruses in nature. Peer J. 9, 1-23. https://doi.org/10.7717/peerj.11088

Zhang YZ, Chen YM, Wang W, Qin XC, Holmes EC (2019). Expanding the RNA Virosphere by Unbiased Metagenomics. Annu. Rev. of Virol. *6*, 119–139. https://doi.org/10.1146/annurev-virology-092818-015851

Table 1. Viral isolates used to construct mock viral communities of varying complexity. The taxonomic status of the various viruses is indicated, together with their DSMZ catalogue code, their propagation host and the GenBank accession number(s) of their genomic sequence(s).

Family	**Genus**	**Virus species**	**Acronym**	**Genome**	**Code[a]**	**Host[b]**	**Sequence accession number(s)**
Alphaflexiviridae	*Allexivirus*	*Shallot virus X*	ShVX	ssRNA(+)			MW854280
Alphaflexiviridae	*Potexvirus*	*Lettuce virus X*	LeVX	ssRNA(+)			MW248356
Benyviridae	*Benyvirus*	*Beet necrotic yellow vein virus*	BNYVV	ssRNA(+)			OK181765-67; M36896
Betaflexiviridae	*Capillovirus*	*Apple stem grooving virus*	ASGV	ssRNA(+)			MW582790
Betaflexiviridae	*Carlavirus*	*Poplar mosaic virus*	PopMV	ssRNA(+)			ON924213
Betaflexiviridae	*Trichovirus*	*Apple chlorotic leaf spot virus*	ACLSV	ssRNA(+)			OK340218; OK340219[c]
Betaflexiviridae	*Tepovirus*	*Potato virus T*	PVT	ssRNA(+)			MZ405665
Bromoviridae	*Alfamovirus*	*Alfalfa mosaic virus*	AMV	ssRNA(+)			MZ405653-55
Bromoviridae	*Anulavirus*	*Pelargonium zonate spot virus*	PZSV	ssRNA(+)			ON398493-95
Bromoviridae	*Bromovirus*	*Brome mosaic virus*	BMV	ssRNA(+)			MW582787-89
Bromoviridae	*Cucumovirus*	*Peanut stunt virus*	PSV	ssRNA(+)			MW307259-61
Bromoviridae	*Ilarvirus*	*Parietaria mottle virus*	PMoV	ssRNA(+)			MZ405646-48
Closteroviridae	*Closterovirus*	*Beet yellows virus*	BYV	ssRNA(+)			MT815988
Closteroviridae	*Crinivirus*	*Tomato chlorosis virus*	ToCV	ssRNA(+)			ON398512-13
Potyviridae	*Bymovirus*	*Barley yellow mosaic virus*	BaYMV	ssRNA(+)			OL311692-93
Potyviridae	*Ipomovirus*	*Cucumber vein yellowing virus*	CVYV	ssRNA(+)			OK181771
Potyviridae	*Potyvirus*	*Bidens mottle virus*	BiMoV	ssRNA(+)			ON398504
Potyviridae	*Rymovirus*	*Agropyron mosaic virus*	AgMV	ssRNA(+)			OM471970
Potyviridae	*Tritimovirus*	*Brome streak mosaic virus*	BrSMV	ssRNA(+)			NC_003501
Potyviridae	Unassigned	*Spartina mottle virus*	SpMV	ssRNA(+)			MN788417

Secoviridae	*Cheravirus*	*Arracacha virus B*	AVB	ssRNA(+)			MW582785-86
Secoviridae	*Comovirus*	*Squash mosaic virus*	SqMV	ssRNA(+)			ON398498-99
Secoviridae	*Fabavirus*	*Broad been wilt virus 1*	BBWV-1	ssRNA(+)			MT663310-11
Secoviridae	*Nepovirus*	*Tomato black ring virus*	TBRV	ssRNA(+)			MW057704-05
Secoviridae	*Sequivirus*	*Carrot necrotic dieback virus*	CNDV	ssRNA(+)			MW080951
Secoviridae	*Stralirivirus*	*Strawberry latent ringspot virus*	SLRSV	ssRNA(+)			MZ405640-41
Solemoviridae	*Sobemovirus*	*Rice yellow mottle virus*	RYMV	ssRNA(+)			MT701719
Solemoviridae	*Enamovirus*	*Pea enation mosaic virus 1*	PEMV1	ssRNA (+)			MW961146
Solemoviridae	*Polerovirus*	*Cucurbit aphid-borne yellows virus*	CABYV	ssRNA(+)			MZ202344
Tombusviridae	*Alphacarmovirus*	*Calibrachoa mottle Virus*	CbMV	ssRNA(+)			OK181769
Tombusviridae	*Alphanecrovirus*	*Tobacco necrosis virus A*	TNV-A	ssRNA(+)			MT675968
Tombusviridae	*Aureusvirus*	*Johnsongrass chlorotic stripe mosaic virus*	JCSMV	ssRNA(+)			MT682309
Tombusviridae	*Betacarmovirus*	*Turnip crinkle virus*	TCV	ssRNA(+)			OK181761
Tombusviridae	*Betanecrovirus*	*Beet black scorch virus*	BBSV	ssRNA(+)			OK058516
Tombusviridae	*Dianthovirus*	*Carnation ringspot virus*	CRSV	ssRNA(+)			MT682300-01
Tombusviridae	*Gammacarmovirus*	*Melon necrotic spot virus*	MNSV	ssRNA(+)			ON398496
Tombusviridae	*Machlomovirus*	*Maize chlorotic mottle virus*	MCMV	ssRNA(+)			OK181780
Tombusviridae	*Pelarspovirus*	*Pelargonium line pattern virus*	PLPV	ssRNA(+)			MW854266
Tombusviridae	*Tombusvirus*	*Tomato bushy stunt virus*	TBSV	ssRNA(+)			MW582792
Tombusviridae	*Umbravirus*	*Carrot mottle virus*	CMoV	ssRNA(+)			OK058520
Tombusviridae	*Umbravirus*	*Pea enation mosaic virus 2*	PEMV2	ssRNA(+)			MW961147; MW961148[c]
Tospoviridae	*Orthotospovirus*	*Impatiens necrotic spot virus*	INSV	ssRNA(+/-)			MW582795-97
Tymoviridae	*Tymovirus*	*Turnip yellow mosaic virus*	TYMV	ssRNA(+)			ON924209
Virgaviridae	*Furovirus*	*Soil-borne wheat mosaic virus*	SBWMV	ssRNA(+)			MZ405651-52
Virgaviridae	*Hordeivirus*	*Barley stripe mosaic virus*	BSMV	ssRNA(+)			ON92421012
Virgaviridae	*Pecluvirus*	*Peanut clump virus*	PCV	ssRNA(+)			NC_003668; NC_003672

Virgaviridae	*Pomovirus*	*Potato mop-top virus*	PMTV	ssRNA(+)			ON398500-02
Virgaviridae	*Tobamovirus*	*Paprika mild mottle virus*	PaMMV	ssRNA(+)			OK181768
Virgaviridae	*Tobravirus*	*Pea early-browning virus*	PEBV	ssRNA(+)			MW854268-69
not assigned	*Idaeovirus*	*Raspberry bushy dwarf virus*	RBDV	ssRNA(+)			MW582777-78
Rhabdoviridae	*Cytorhabdovirus*	*Lettuce necrotic yellows virus*	LNYV	ssRNA(-)			MZ202327
Rhabdoviridae	*Varicosavirus*	*Beet oak leaf virus*	BOLV	ssRNA(-)			OK181765-67
Rhabdoviridae	*Alphanucleorhabdovirus*	*Physostegia chlorotic mottle virus*	PhCMoV	ssRNA(-)			KX636164
Rhabdoviridae	*Betanucleorhabdovirus*	*Sonchus yellow net virus*	SYNV	ssRNA(-)			MT613317
Aspiviridae	*Ophiovirus*	*Lettuce ring necrosis virus*	LRNV	ssRNA(-)			ON398506-09
Partitiviridae	*Alphacryptovirus*	*Poinsettia latent virus*	PnLV	dsRNA			ON398503
Caulimoviridae	*Badnavirus*	*Banana streak OL virus*	BSOLV	dsDNA-RT			NC_003381
Caulimoviridae	*Caulimovirus*	*Cauliflower mosaic virus*	CaMV	dsDNA-RT			NC_001497
Geminiviridae	*Begomovirus*	*Squash leaf curl virus*	SLCV	ssDNA			MW582809-10
Geminiviridae	*Mastrevirus*	*Maize streak virus*	MSV	ssDNA			NC_001346
Nanoviridae	*Babuvirus*	*Banana bunchy top virus*	BBTV	ssDNA			NC_003473-77; NC_003479

(a) DSMZ catalogue code (information expected from DSMZ collaborators)

(b) Host in which the virus isolate was propagated and lyophilized (information expected from DSMZ collaborators)

(c) Two different variants are present in the propagated sample and accession numbers for both variants are provided

Table 2. Additional viruses identified by analysis of the HTS data in the samples used to assemble the synthetic mock communities of varying complexity.

Family	Genus	Virus species	Acronym	Genome type	Reference sequence accession number[a]
Virgaviridae	*Tobamovirus*	*Tobacco mosaic virus*	TMV	ssRNA(+)	Acc. number expected
Tymoviridae	unassigned	*Poinsettia mosaic virus*	PnMV	ssRNA(+)	Acc. number expected
Endornaviridae	*Alphaendornavirus*	*Hordeum vulgare endornavirus*	HvEV	ssRNA(+)	Acc. number expected
Solemoviridae	*Polerovirus*	*Turnip yellows virus*	TuYV	ssRNA(+)	JQ862472
Geminiviridae	*Mastrevirus*	*Maize streak Réunion virus*	MSRV	ssDNA	Acc. number expected
Totiviridae	unassigned	*Maize-associated totivirus*	MATV	dsRNA	Acc. number expected
Totiviridae	unassigned	*Maize-associated totivirus 2*	MTV-2	dsRNA	MN428829
Mitoviridae	*Duamitovirus*	*Chenopodium quinoa mitovirus 1*	CqMV1	ssRNA(+)	MT089917
small linear ssRNA satellite		*Turnip crinkle satellite RNA F*	TCVsatRNA F	ssRNA	X12749
small linear ssRNA satellite		*Pea enation mosaic virus satellite RNA*	PEMVsatRNA	ssRNA	Acc. number expected
small linear ssRNA satellite		*Strawberry latent ringspot virus satellite RNA*	SLRSVsatRNA	ssRNA	Acc. number expected

(a) Accession number of the closest sequence in GenBank that was used as reference for reads mapping

Table 3. Comparison of the number and average length of *de novo* assembled viral contigs obtained for VANA and dsRNA datasets normalized at different sequencing depths (100K, 300K, 1M, 3M and 10M reads, 5 resampling repeats at each depth). The standard deviations (SD) and the statistical differences (p-values) are also shown

		VANA average +/- SD	**dsRNA average +/- SD**	**Two sample t-test**
100 K reads	nb viral contigs	33.6 +/- 1.9	101.8 +/- 2.9	*9.2E-11*
	Viral contigs average length	733.4 +/- 23.7	747.4 +/- 17.1	0.32
300K reads	nb viral contigs	70.2 +/- 5.4	129.4 +/- 8.1	*8.0E-07*
	Viral contigs average length	643.4 +/- 27.8	887.8 +/- 38.2	*2.8E-06*
1M reads	nb viral contigs	106.2 +/- 6.3	159.2 +/- 6.6	*1.1E-06*
	Viral contigs average length	694.8 +/- 30.3	1019.6 +/- 40.9	*5.7E-07*
3M reads	nb viral contigs	129.6 +/- 4.8	207.6 +/- 3.8	*2.5E-09*
	Viral contigs average length	798.4 +/- 15.9	1067.6 +/- 11.5	*1.4E-09*
10M reads	nb viral contigs	201.2 +/- 4.1	268 +/- 2.9	*1.8E-09*
	Viral contigs average length	791.2 +/- 11.1	1121.4 +/- 10.6	*3.9E-11*

Table 4. Performance parameters of *de novo* assembly using different minimal contigs length of normalized, 10M reads, VANA and dsRNA datasets for the 60-viruss synthetic community.

	Minimal contig length											
	125 nt		**175 nt**		**250 nt**		**350 nt**		**500 nt**		**1000 nt**	
	VANA	**dsRNA**	**VANA**	**dsRNA**	**VANA**	**dsRNA**	**VANA**	**dsRNA**	**VANA**	**dsRNA**	**VANA**	**dsRNA**
nb contigs	1947	2212	416	784	220	437	144	276	86	182	37	88
average length	235	324	506	607	757	907	985	1243	1355	1662	2191	2696
N25	547	1764	1836	3642	2334	4007	2560	4060	3449	4505	3824	5671
N50	206	352	628	1005	994	1521	1277	1955	1709	2775	2277	3705
N75	156	191	313	352	481	558	618	773	888	1117	1653	1782
Max	6549	13919	6652	13919	6652	13919	6549	13919	6652	13919	6652	13919
nb viral contigs	1852	1672	378	468	204	269	137	181	84	131	37	75
% viral contigs	95%	76%	91%	60%	93%	62%	95%	66%	98%	72%	100%	85%
Viral contigs average length	235	327	525	741	783	1123	1008	1508	1368	1921	2191	2833
Bases in viral contigs	435421	547507	198486	347003	159827	302074	138102	272883	114951	251656	81077	212467
% bases in viral contigs	95.20%	76.40%	94.40%	72.90%	96.00%	76.20%	97.40%	79.50%	98.60%	83.20%	100%	89.50%

Table 5. Detection performance of VANA and dsRNA methods at the level of individual viral genomic molecules (a total of 115 viral molecules) using the most complex, 60-virus pool, for groups of viruses with different genome types at 10M reads sequencing depth.

	# viral molecules	**VANA**		**dsRNA**	
		# detected	**% detected**	**# detected**	**% detected**
+ssRNA viruses	86	50	58.1%	77	89.5%
-ssRNA viruses	12	1	8.3%	5	41.7%
RNA satellites	3	1	33.3%	3	100%
dsRNA viruses	2	0	0%	2	100%
ssDNA viruses	10	4	40.0%	0	0%
dsDNA viruses	2	1	50.0%	1	50.0%
Total	**115**	**57**	**49.6%**	**88**	**76.5%**

Figure 1. Average proportion of viral reads in VANA and dsRNA datasets from viral communities of different complexities (A) and number of RNA viruses detected at an even 120K read depth for communities of different complexities (B).

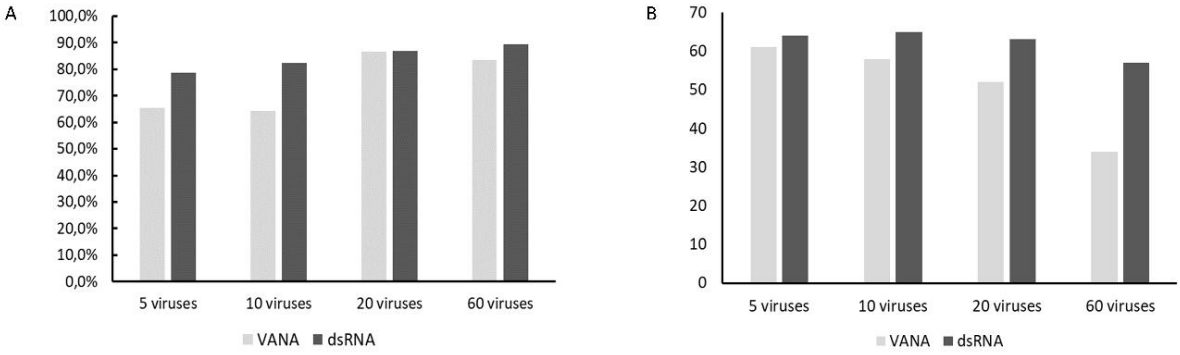


Figure 2: Distribution of percentage of mapped reads for each detected virus in the 60-viruses community using a normalized 1.44 million reads sequencing depth.

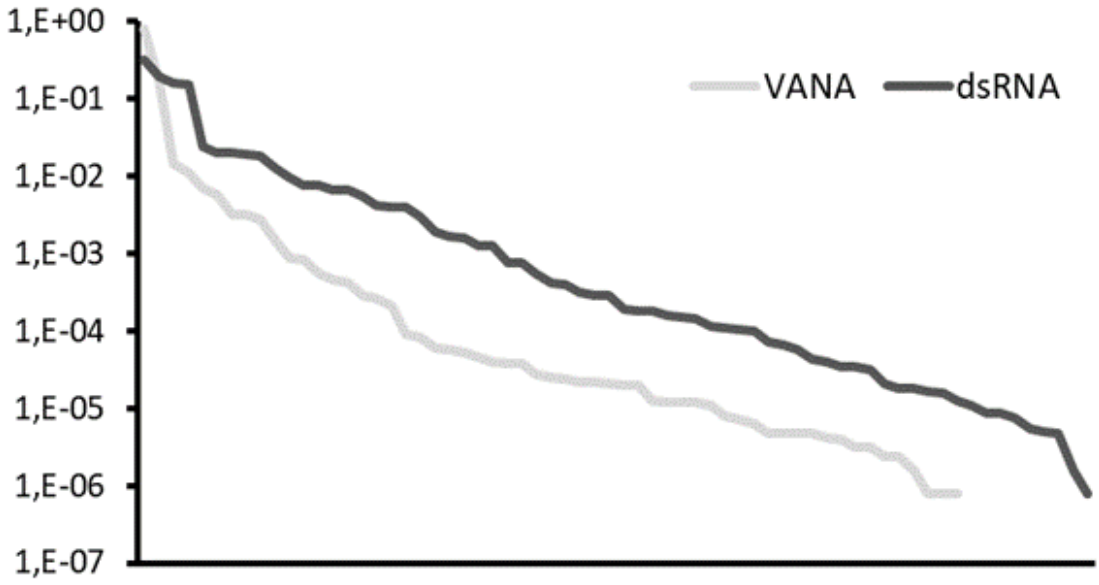


Figure 3. Number of detected RNA viruses in the 60-viruses community as a function of minimal contig length at an equal sequencing depth of 10M reads.

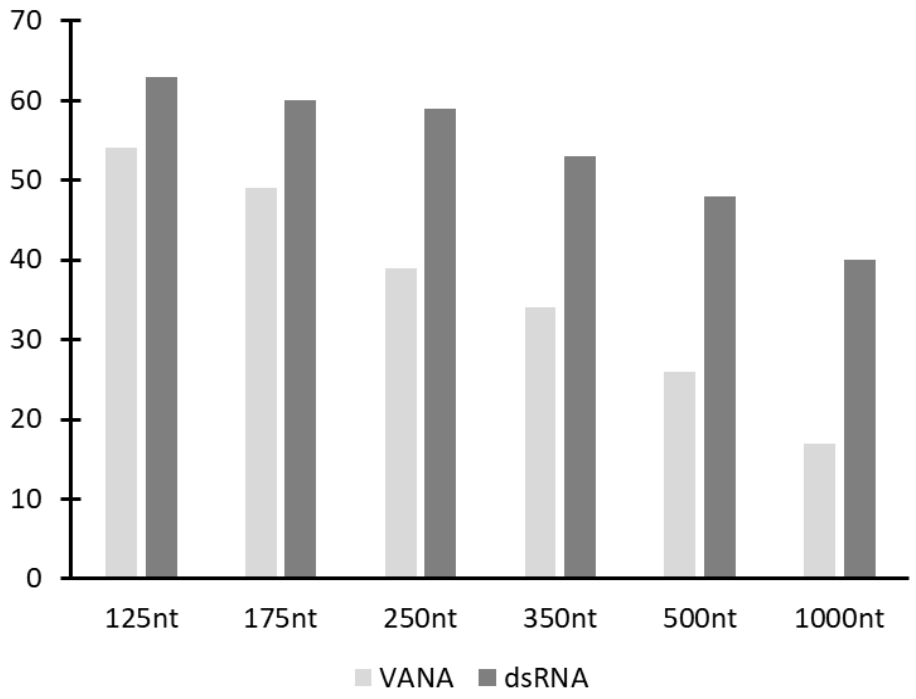
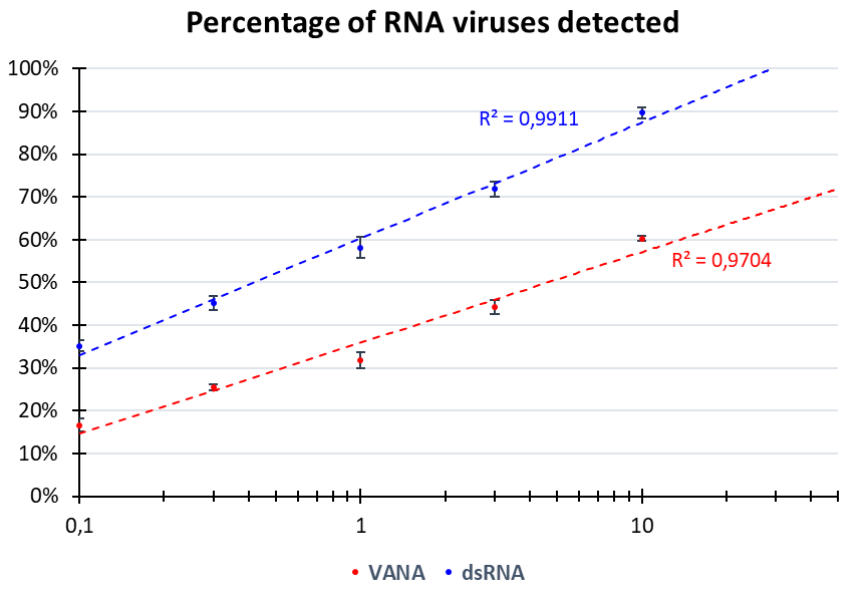


Figure 4: Observed percentages of regression over a logarithmic scale of sequencing depths with the linear regression r^2 value shown for both, dsRNA and VANA approaches.



Supplementary Materials [Supplementary Table S1 that does not fit easily into an A4 format is available at https://doi.org/10.57745/XYIKFK]

Supplementary Tables

Supplementary Table S1. Pooling strategy to generate the various pools of variable complexity (from 5 to 60 viruses in a pool).

Supplementary Table S2. Comparison of *de novo* assembly parameters for VANA and dsRNA datasets normalized at different sequencing depths (100K, 300K, 1M, 3M and 10M reads, 5 resampling repeats at each depth) and corresponding statistical significance.

Supplementary Figures

Supplementary Figure S1: Pooling strategy to generate mock virus communities with different degrees of complexity (5, 10, 20, and 60-viruses communities).

Supplementary Figure S2: Percent coverage of detected viral molecules using the VANA or the dsRNA approaches as a function of minimal contig length.

Supplementary Figure S3: Number of detected RNA viruses based on *de novo* assembled contigs from the VANA or the dsRNA approaches for datasets normalized at a 120K reads sequencing depth and for viral communities with different degrees of complexity.

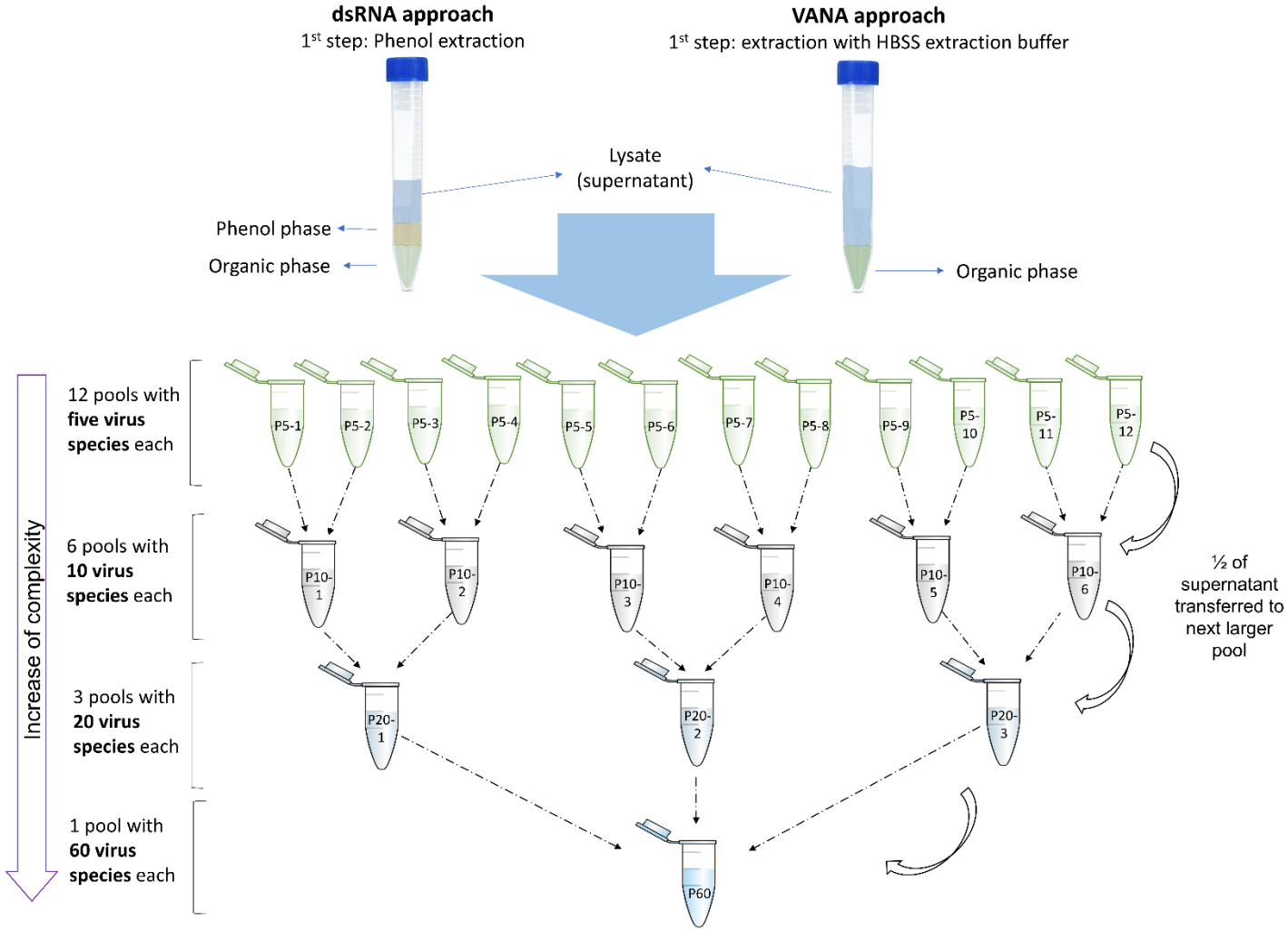
Supplementary Figure S4: Number of detected RNA viruses for viral communities with different degrees of complexity using *de novo* assembled contigs from the VANA or the dsRNA approaches derived from datasets normalized so as to compensate for community complexity (120K reads for 5 viruses communities, 240K for 10 viruses, 480K for 20 viruses and 1.44M for 60 viruses).

Supplementary Figure S5: Average proportion of the length of viral molecules represented by contigs obtained for the VANA or the dsRNA approaches as a function of sequencing depth. For each sequencing depth 5 independent random resamplings were performed and error bars represent the standard deviations of the coverage obtained.

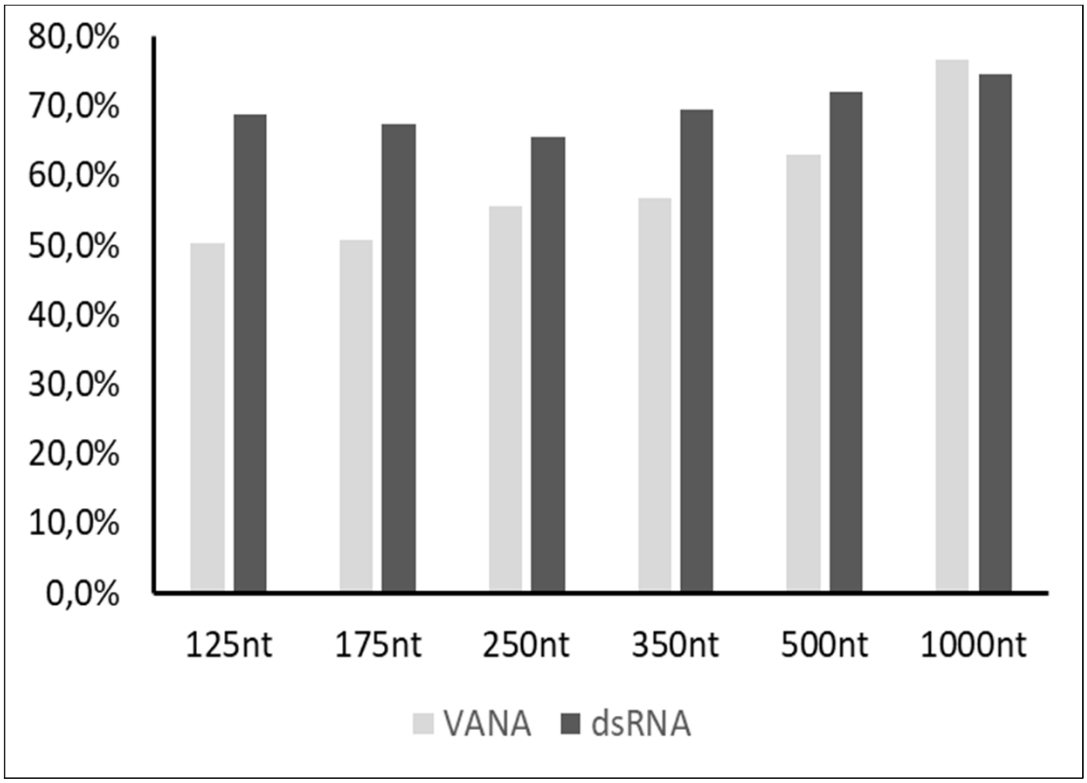
Supplementary Table S2.

		VANA average +/- SD	**dsRNA average +/- SD**	**Two sample t-test**
100K reads	nb contigs	33.6 +/- 1.9	103.2 +/- 2.4	*2.6E-11*
	average length	733.4 +/- 23.7	741.8 +/- 14.9	0.52
	N50	839.2 +/- 104.8	937.2 +/- 34.1	0.10
	Max length	5886.2 +/- 442.4	5277.2 +/- 577.9	0.10
	nb viral contigs	33.6 +/- 1.9	101.8 +/- 2.9	*9.2E-11*
	% viral contigs	100% +/- 0%	99% +/- 1%	*2.5E-02*
	Viral contigs average length	733.4 +/- 23.7	747.4 +/- 17.1	0.32
300K reads	nb contigs	70.6 +/- 6.1	140.4 +/- 8.9	*5.1E-07*
	average length	642.4 +/- 30.0	849.4 +/- 35.3	*8.5E-06*
	N50	681.2 +/- 93.1	1102.4 +/- 104.5	*1.5E-04*
	Max length	6260 +/-213.1	7581.6 +/- 3875.5	0.49
	nb viral contigs	70.2 +/- 5.4	129.4 +/- 8.1	*8.0E-07*
	% viral contigs	100% +/- 1%	92% +/- 2%	*7.7E-05*
	Viral contigs average length	643.4 +/- 27.8	887.8 +/- 38.2	*2.8E-06*
1M reads	nb contigs	108.2 +/- 6.1	198.4 +/- 7.8	*3.5E-08*
	average length	687.6 +/- 28.1	915.4 +/- 35.4	*3.5E-06*
	N50	767 +/- 74.9	1556.4 +/- 178.3	*2.6E-04*
	Max length	6491 +/- 79.4	10382.6 +/- 1825.7	*8.9E-03*
	nb viral contigs	106.2 +/- 6.3	159.2 +/- 6.6	*1.1E-06*
	% viral contigs	98% +/- 1%	80% +/- 2%	*8.3E-08*
	Viral contigs average length	694.8 +/- 30.3	1019.6 +/- 40.9	*5.7E-07*
3M reads	nb contigs	134.2 +/- 3.4	284 +/- 8.2	*2.5E-07*
	average length	783.4 +/- 10.1	931 +/- 16.6	*1.5E-07*
	N50	1016.6 +/- 73.5	1599.4 +/- 214.1	*4.3E-04*
	Max length	6540 +/- 3.4	11573.4 +/- 2359.2	*8.8E-03*
	nb viral contigs	129.6 +/- 4.8	207.6 +/- 3.8	*2.5E-09*
	% viral contigs	97% +/- 1%	73% +/- 1%	*1.0E-09*
	Viral contigs average length	798.4 +/- 15.9	1067.6 +/- 11.5	*1.4E-09*
10M reads	nb contigs	217 +/- 4.4	433.8 +/- 4.4	*8.2E-13*
	average length	764.6 +/- 10.5	907.2 +/- 7.7	*8.2E-09*
	N50	1025.4 +/- 31.2	1529.6 +/- 21.7	*1.8E-09*
	Max length	6534.4 +/- 210.2	13930.4 +/- 24.9	*1.6E-07*
	nb viral contigs	201.2 +/- 4.1	268 +/- 2.9	*1.8E-09*
	% viral contigs	93% +/- 0%	62% +/- 1%	*2.2E-13*
	Viral contigs average length	791.2 +/- 11.1	1121.4 +/- 10.6	*3.9E-11*

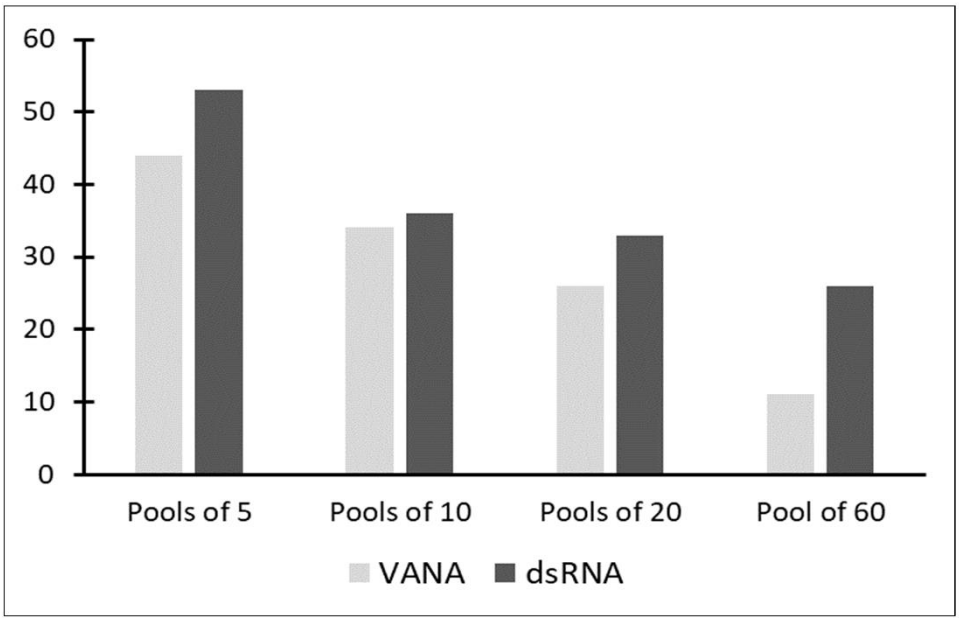
Supplementary Figure S1



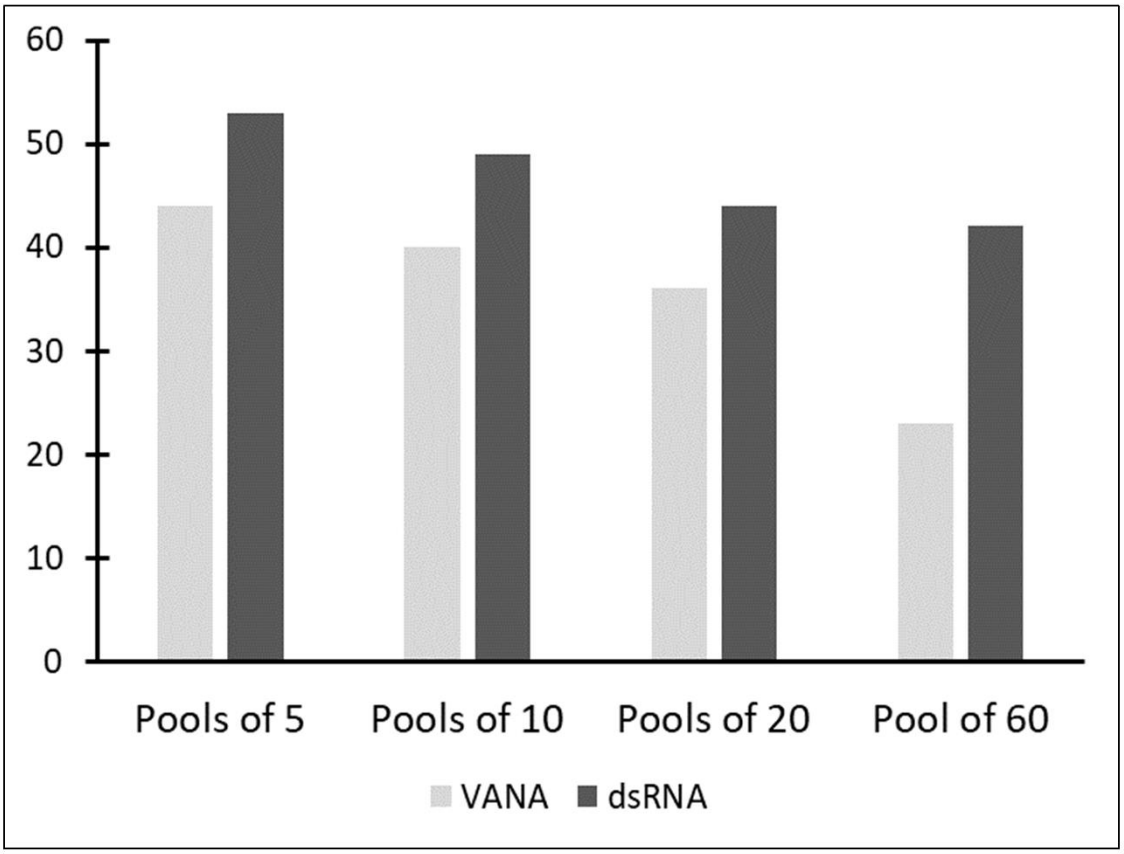
Supplementary Figure S2



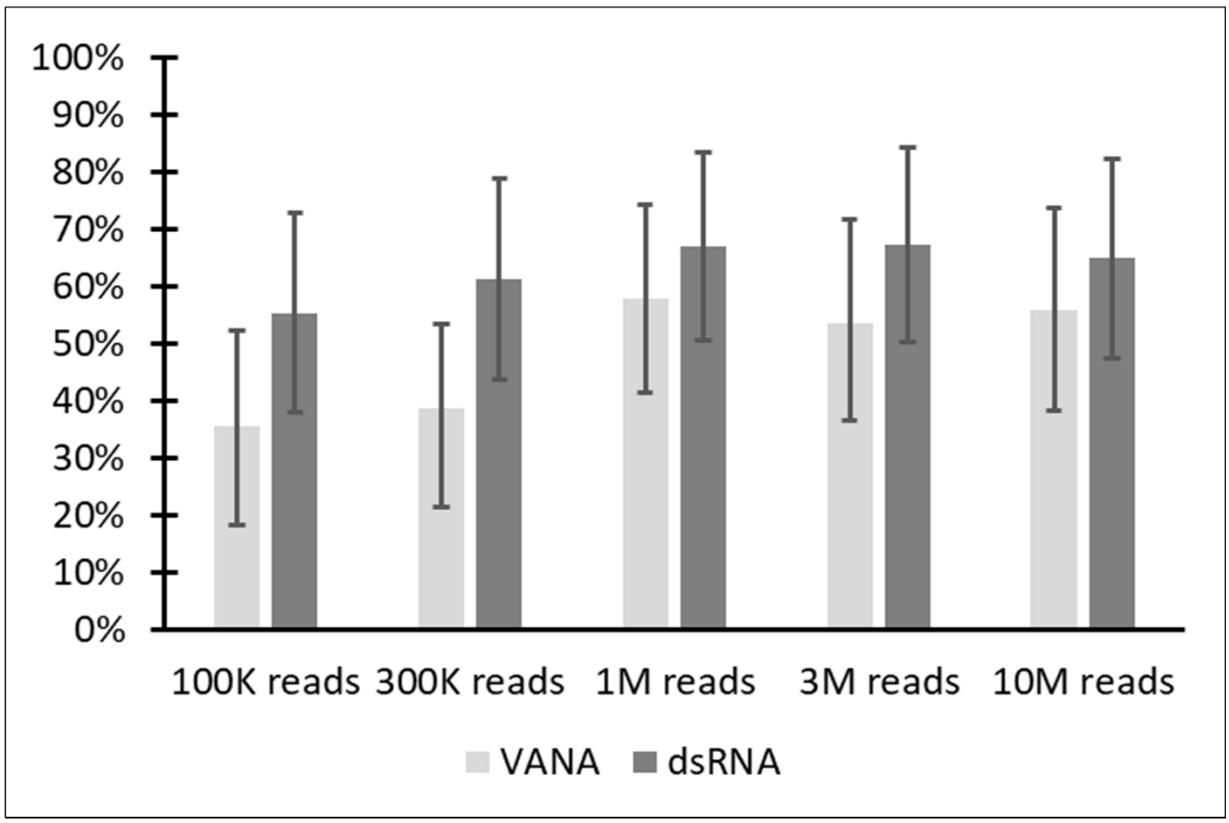
Supplementary Figure S3



Supplementary Figure S4



Supplementary Figure S5



CHAPTER II

Carrot populations in France and Spain host a complex virome rich in previously uncharacterized viruses

This chapter is divided into three manuscripts, the first (to be submitted shortly) describing all novel and known viruses detected in the different carrot populations sampled in France and in Spain, and two recently published articles, one describing a novel flavi-like virus in wild carrots and the other discussing the diversity of polerovirus-associated RNAs identified in the carrot virome.

Carrot populations in France and Spain host a complex virome rich in previously uncharacterized viruses

Deborah Schönegger[1], Armelle Marais[1], Bisola Babalola[2], Chantal Faure[1], Marie Lefebvre[1], Laurence Svanella-Dumas[1], Sára Brázdová[1$], Thierry Candresse[1*]

[1] Univ. Bordeaux, INRAE, UMR BFP, CS20032, 33882, Villenave d'Ornon Cedex, France.

[2] Centro de Biotecnología y Genómica de Plantas (CBGP), Universidad Politécnica de Madrid (UPM) and E.T.S.I. Agronómica, Alimentaria y de Biosistemas, Campus de Montegancedo, UPM, 28223 Pozuelo de Alarcón, Madrid, Spain.

$ **Present address**: Department of Plant Virology, Institute of Plant Molecular Biology, Biology Centre of the Czech Academy of Sciences, Branišovská 31, České Budějovice, Czech Republic

*** Author for correspondence:** thierry.candresse@inrae.fr Tel : (+33) 557 12 23 89

Abstract word count: 222 words

Text word count: 9287 words (Introduction-Materials and Methods-Results-Discussion)

Tables and Figures: 2 Tables, 11 Figures, 14 supplementary Tables and Figures

Running title: Analysis of the carrot virome

The nucleotide sequences reported in this work have been deposited in the GenBank database under accession numbers OP886449 to OP886481 and OP889248 to OP889249

Abstract

High-throughput sequencing (HTS) has been proven a powerful tool to uncover the virome of cultivated and wild plants and offers the opportunity to study virus movements across the agroecological interface. The carrot model consisting of cultivated (*Daucus carota* ssp. *sativus*) and wild carrot (*Daucus carota* ssp. *carota*) populations, is particularly interesting with respect to comparisons of virus communities due to the low genetic barrier to virus flow since both types belong to the same plant species. Using a highly purified double-stranded RNA-based HTS approach, we analyzed on a large scale the virome of 45 carrot populations including cultivated, wild and off-type carrots (wild carrots growing within the field and likely representing hybrids between cultivated and wild carrots) in France and additional six carrot populations in central Spain. Globally, we identified a very rich virome including 45 viruses of which 25 are novel. Most of the identified novel viruses showed preferential associations with wild carrots, either occurring exclusively in wild populations or infecting only a small proportion of cultivated populations, indicating the role of wild carrots as a reservoir of viral diversity. In addition, the carrot virome is particularly rich in viruses involved in complex interactions of mutual interdependencies for aphid transmission such as poleroviruses, umbraviruses and associated satellites which can be the basis for further investigations of synergistic or antagonistic virus-vector-host relationships.

Keywords: carrot, Daucus carota, virome, double-stranded RNA, high-throughput sequencing

INRTODUCTION

Agriculture has drastically changed biodiversity at the landscape level by converting natural habitats into intensively managed environments. As a result, an increasing proportion of natural habitats lie adjacent to agricultural lands, referred to as the agroecological interface, which integrates wild lands as highly fragmented patches within a matrix of crop-, pasture- and grassland (Alexander et al., 2014). Such intersections of managed and unmanaged ecosystems can directly influence the emergence of pathogens via pathogen spill-over and spill-back events between them (Alexander et al., 2014; Elena et al., 2014; Roossinck and García-Arenal, 2015). This is particularly important with regard to viruses that can infect both native wild plants and cultivated crops, allowing them to move through the ecological interface (Roossinck and García-Arenal, 2015). Wild plants can therefore play a role as reservoirs of virus infecting cultivated plants. On the other hand, viruses infecting cultivated plants can spread into wild plants and influence their fitness within unmanaged ecosystems in more or less complex/subtle ways (Malmstrom and Alexander, 2016). The agroecological interface is therefore an important boundary that can contribute to the structuring of virus communities and thus directly and indirectly influence ecosystem functions. Viruses are ubiquitous biological entities in diverse environments and play a significant role as pathogens in agriculture, where they are estimated to be responsible for half of emerging diseases (Anderson et al., 2004). However, in some instances and in particular when environmental conditions change, viruses have been suggested to switch from parasitic to mutualistic interactions with their host (González et al., 2021). In addition, in wild plants, viruses are often asymptomatic and it has been suggested that one of the reasons for this situation might be viruses long co-evolution with wild hosts and vectors in complex trophic interactions (Lefeuvre et al., 2019).

The family *Apiaceae* comprises about 3,700 species consisting of wild plants and important crops (Petruzzello, 2019). Carrot (*Daucus carota* L) is the most widely grown crop within the family. In total, more than 30 viruses from 12 different families are known to infect carrots (Mihara et al., 2016; Brunt et al., 1996; Wu et al., 2022), with some of them causing severe damage to carrot production. Early studies have focused on carrot necrotic dieback virus (CNDBV, genus *Sequivirus*) and the carrot motley dwarf (CMD) disease complex. CNDBV was originally described as the Anthriscus strain of parsnip

yellow fleck virus (PYFV, genus *Sequivirus*) by van Dijk and Bos (1985) as it naturally infects cow parsley (*Anthriscus sylvestris* L.). In carrots, it causes leaf necrosis, severe stunting and dieback (Menzel and Vetten, 2008). The CMD complex was first described in Australia by Stubbs (1948) and is caused by a mixed infection of a polerovirus, carrot red leaf virus (CtRLV), an umbravirus, carrot mottle virus (CMoV) or carrot mottle mimic virus (CMoMV) and, in some instances, a polerovirus-associated RNA, carrot red leaf virus-associated RNA (CtRLVaRNA) (Yoshida, 2020). The umbravirus and the associated RNA lack a capsid protein and their RNAs are transcapsidated by the coinfecting polerovirus for transmission by the aphid *Cavariella aegopodii*. The CMD complex causes leaf reddening and yellowing, stunting and reduces sugar content of taproots (Yoshida, 2020; Stubbs, 1948). Another economically important virus causing disease in carrot crops is carrot yellow leaf virus (CYLV), a closterovirus, which has been associated with internal root necrosis (Adams et al., 2014). The authors found additional novel viruses infecting carrot crops, such as novel members of the *Chordovirus* and *Closterovirus* genera. However, less is known about viruses infecting wild carrots which may represent an important reservoir of carrot infecting viruses.

Recent advances in sequencing technologies such as high-throughput sequencing (HTS) have revolutionized the field of viral ecology over the past decade by moving from a "one host-one pathogen" perspective to analyzing the entire phytovirome at ecosystem level. HTS-based studies have highlighted the importance of wild plants as a source of virus diversity (Muthukumar et al., 2009; Roossinck, 2012, Thapa et al., 2015; Bernardo et al., 2018; Susi et al., 2019). More recent studies have confirmed the importance of unmanaged wild plants surrounding cultivated crop species in the epidemiology, ecology, distribution and emergence of viruses (Bernardo et al., 2018; Ma et al., 2020; Ma et al., 2021; Maachi et al., 2022, Rivarez et al., 2022). Carrot represents a particularly interesting model plant as the cultivated crop (*Daucus carota* ssp. *sativus*) and its weedy relative, wild carrot (*Daucus carota* ssp. *carota*), belong to the same species and therefore represent a host pair with the lowest expected host genetic barrier to virus flow. The aim of the present study was to characterize virus communities in wild and cultivated carrots using a double-stranded RNA (dsRNA)-based metagenomic approach. We have identified a rich and diverse carrot virome in different regions of France and central Spain, providing the basis for further

studies on the role of wild plants as virus reservoirs and on the complex interplay and interdependencies of viruses in mixed infections as well as the underlying evolutionary processes in relation to their vectors and hosts.

MATERIALS AND METHODS

Sample collection and processing

In July-August 2019, 16 different carrot populations were sampled in south-west France (Gironde and Landes departments of the Nouvelle Aquitaine region) (Supplementary Table S1). The sampled carrot populations comprised four cultivated populations, four populations growing within cultivated carrot fields and showing some characteristics of wild carrots (early bolting, small tap root, often white in color) suggesting an origin from seed contaminations of the planted commercial varieties and thereafter referred to as off-type carrots, and eight wild carrot populations. In July-September 2020, carrot populations were sampled across a north south transect of France (Supplementary Table S1). In total, 29 carrot populations were collected, comprising 11 cultivated carrots populations, three off-type populations and 15 wild populations. A total of 45 carrot populations were therefore sampled in France over the two years (Supplementary Table S1, Figure 1A). Additional carrot populations were sampled in June 2021 near Segovia (central Spain), including five cultivated and one wild carrot populations (Supplementary Table S1, Figure 1B).

In each sampled population, leaves from 50 individual plants were collected. Carrot plants were sampled regardless of the presence of viral symptoms, but plants with evident fungal infection or necrosis were excluded. Leaf samples were stored desiccated over anhydrous $CaCl_2$ (Sigma-Aldrich, Saint-Quentin-Fallavier, France) until use. For each population, a pool corresponding to the 50 sampled plants was assembled (0.1 g leaf per plant).

Double-stranded RNAs purification and Illumina sequencing

Double-stranded (ds) RNAs were purified from each pool consisting of 50 carrot plants. Samples were ground to a fine powder using liquid nitrogen and a mortar and pestle and dsRNAs purified by two rounds of CC41 cellulose chromatography as described by Marais et al. (2018). A negative extraction control was included by using only buffer. In some cases, dsRNAs were also extracted from individual

plants using the same protocol. Purified dsRNAs were converted to cDNA and randomly amplified while simultaneously adding MID tags by RT-PCR (François et al., 2018; Marais et al., 2018). Random amplification PCR products were purified using the MinElute PCR purification kit (Qiagen, Courtaboeuf, France) and their concentration determined spectrophotometrically. Equal amounts of amplification products from each sample/pool were sent for Illumina sequencing in a multiplexed format (2×150 bp) on the NovaSeq 6000 system at the GetPlaGe platform (GenoToul INRAE Toulouse, France).

Bioinformatic analysis of Illumina sequencing reads: demultiplexing, quality trimming, *de novo* assembly, contigs annotation and read-mapping

Sequencing reads were demultiplexed and trimmed on quality using CLC Genomic Workbench, version 21.0.3 (CLC-GW, Qiagen) using default settings and a minimum read length of 60 nucleotides (nt). Contigs were assembled *de novo* using a minimal contig length of 250 nt. Viral contigs were annotated by a comparison against the non-redundant sequence (nr) database in GenBank using BlastX (Altschul et al., 1990). Identified contigs were manually scaffolded if needed using the most closely related reference sequence from NCBI GenBank. They were verified and extended by repeated rounds of mapping of residual reads using CLC-GW. When a given virus was identified in multiple samples, the longest contig with the deepest coverage was finally selected to be used as a reference sequence for that virus and was deposited in GenBank. Novel virus species were identified and tentative taxonomic assignation proposed based on the calculation of percent pairwise identities calculated using Mega7 (Kumar et al., 2018) and multiple nucleotide or amino acid alignments with reference sequences selected according to the International Committee on Taxonomy of Viruses (ICTV) species demarcation criteria for each virus genus or family.

In order to determine the geographic distribution of the identified viruses, reads from datasets, normalized to a common depth per sample using the read sampling tool in CLC-GW, were mapped back to reference contigs (for novel viruses) and reference genomic sequences from GenBank (for known viruses) using CLC-GW and nt identity and length fraction set at a minimum of 90%. The mapped reads were then manually checked to ensure the identity of the detected virus. In order to address the well-

known HTS inter-sample cross talk issue, two approaches were used to estimate it. The first one considered the viral reads identified for each virus in the blank control for each run and was determined as average background reads plus 3 standard deviations. The second one used data from individually analysed plants for which we had both RT-PCR and HTS data so that plants that tested negative in RT-PCR but showed a few reads of the given virus in the corresponding dataset could be used calculate an average number of background reads per million sequencing reads. With this strategy, a separate detection threshold could be calculated for each of those viruses that were screened by RT-PCR. For viruses not tested by RT-PCR, we used as threshold the average of the threshold values determined for the RT-PCR tested viruses. Finally, the thresholds obtained for each virus using either the blank controls or the RT-PCR testing were compared and the highest value retained to establish the final HTS positive detection threshold for each virus and clean the reads mapping table for each sample/virus combination.

In some cases, genomes were completed by determining 5' and 3' genome ends by Rapid Amplification cDNA Ends (RACE) experiments using specific primers (Supplementary Table S2) and the SMARTer RACE Kit (Takara Bio Europe, Saint-Germain-en-Laye, France), as recommended by the company.

Specific two-step RT-PCR assays (Marais et al., 2011) were developed for selected novel viruses to identify infected individual plants from positive pools. Primers were designed manually in MEGA 7 using multiple alignments generated from contigs of the corresponding viruses and of related viruses and GenBank reference sequences, with the aim to design primers that would amplify all isolates of a given virus but be specific enough not to amplify isolates of closely related viruses (Supplementary Table S2). PCR products were visualized on 1.2 % agarose gels and selected amplicons were directly Sanger sequenced (Eurofins Genomics, France) to confirm the specificity of the amplification products.

Analysis of viral genomic sequences, phylogenetic and recombination analyses.

Open reading frames (ORFs) were predicted using CLC-GW and confirmed using the ExPASy Translate Tool (Artimo et al., 2012). Multiple nt or aa sequences alignments were performed using ClustalW as implemented in MEGA 7 (Thompson et al., 1994). Recombination events were analyzed from such multiple alignments using the RDP4 package (Martin et al., 2015). Only recombination events detected by at least four out of seven implemented algorithms were considered. Neighbor-joining trees were

inferred from multiple alignments of whole genomes or conserved gene products and strict nt or aa identities calculated using MEGA 7. The significance of branches was estimated with 1,000 bootstrap replicates.

RESULTS

During summer 2019 and 2020, a total of 45 carrot populations were sampled in France, covering a north-south country gradient. Additionally, five cultivated and one wild carrot populations were sampled near Segovia (central Spain) in 2021. Details on the sampled carrot populations are shown in Supplementary Table S1 and in Figures 1A and 1B. Illumina sequencing of cDNAs prepared from highly purified dsRNAs prepared from pools of 50 plants for each population and subsequent bioinformatic analyses revealed a rich and diverse virome comprising 45 viruses belonging to 12 different viral families. In most cases, long contigs representing near complete genomes with high average coverage could be assembled, paralleling other virome studies, demonstrating the feasibility of assembling long contigs from pools of plants from the same species/location (Ma et al., 2020; Rivarez et al., 2022). Out of these, 20 viruses (44%) belong to already known species, while the remaining 25 viruses (56%) belong to novel species in nine different viral families, or are unclassified subviral agents or putative novel *Riboviria* whose precise higher order taxonomic status remains to be established (Figure 1C). Details of novel and known viruses identified are presented in Tables 1 and 2, respectively. A more detailed description and analysis of the identified viruses is provided in what follows.

Discovery of different known and novel umbraviruses and of a potential novel class of umbra-like virus

The French and Spanish carrot viromes were found to be rich in viruses of the genus *Umbravirus*. All three umbraviruses known to infect carrots, including carrot mottle virus (CMoV), carrot mottle mimic virus (CMoMV) and the more recently reported wild carrot mottle virus (WCMoV)were identified in the three carrot population types (cultivated, off-type and wild), being detected in more than 90% of all sampled populations in France and Spain. In addition, we identified in different cultivated and wild populations from different regions in France a virus of uncertain taxonomic status related to CMoV and

referred to here as carrot umbravirus 1 (CaUV-1), and two novel umbraviruses, carrot umbravirus 2 (CaUV-2) and carrot umbravirus 3 (CaUV-3). In all cases, the genome organization was typical for the genus with four ORFs coding respectively from 5' to 3' for a replication-associated protein, a RNA-dependent RNA polymerase (RdRp, expressed as a fusion as a consequence of a -1 ribosomal frameshift) and long distance movement and cell to cell movement proteins encoded by overlapping 3' ORFs. A phylogenetic tree based on a multiple nt alignment of the RdRp genes for known umbraviruses and for those novel ones reported here is shown in Figure 2A.

A 3.3kb scaffold with >3,000-fold coverage could be obtained for CaUV1, covering an estimated 79% of the full genome. This sequence shows a pairwise nt identity of 78% and 76% to pastinaca umbravirus 1 (PasUV1, GenBank OL472236) and CMoV (GenBank KF533714), respectively, in its partial RdRp gene (Supplementary Table S3), which is above the ICTV species demarcation threshold of 70% nt identity between virus species for the *Umbravirus* genus. However, the available CaUV1 sequence covers only 67% of the full RdRp gene and it is therefore not possible to estimate the pairwise identity that would be reached if full genomes were compared and, therefore, to decide whether CaUV1 should be regarded as a variant of PasUV1 or CMoV or as belonging to a new umbravirus species. A recombination analysis performed using the multiple nt sequence alignment used for phylogenetic analysis suggested the existence of a recombination event linking PasUV1, CMoV and CaUV1. Details of this recombination event, detected by all algorithms and with a combined Bonferoni corrected probability of $1.7xE^{-35}$ are shown in Supplementary Figure S1. This event involves the 3' half of the movement protein (MP) gene and the 3' non-coding region (3'NCR), with up to 98% nt identity in this ca. 1 kb region between CMoV (KF533714) and CaUV1. It is however not possible to decide whether CaUV1 or CMoV should be considered as the parent or the recombinant, since it is not possible to orient this specific recombination event. CaUV1 was detected in all three carrot population types, even though the proportion of populations with CaUV1 infection was slightly higher in the cultivated (8/15, 53.3%) and off-type (4/7, 57.1%) as compared to wild populations (7/23, 30.1%).

Nearly complete genomes of the other two novel umbraviruses, CaUV2 and CaUV3 could be assembled of respectively 4.25 and 4.06 kb and 605x and 1421x average coverage. These assembled genomes show

highest nt pairwise identities of respectively 64% and 61% with carrot mottle mimic virus (CMoMV, GenBank MW848525) and red clover umbravirus (RCUV, GenBank MG596237), based on near complete genome alignments (Supplementary Table S3) and only 47% nt identity with each other. These identity values are clearly below the 70% nt identity cut-off threshold in the *Umbravirus* genus, supporting the notion the sequenced isolates belong to distinct novel species. Both viruses were detected in more than half of the wild populations over the two sampling years in France (65.21%, 15/23) and a smaller proportion of field (cultivated plus off-type) carrot populations. CaUV2 was detected in 33.3% (5/15) of cultivated and 42.8% (3/7) of off-type populations, while the corresponding values for CaUV3 were respectively 13.3% (2/15) and 57.1% (4/7). These two novel umbraviruses were not detected in any of the six Spanish populations.

A fourth partial umbra-like sequence was identified in a total of seven French carrot populations: one wild and one off-type 2019 populations (FR19-2 and FR19-3, Supplementary Table S1) and two cultivated, one off-type and two wild 2020 populations (FR20-2, FR20-4, FR20-6, FR20-15 and FR20-25, Supplementary Table S1). It was also detected in a cultivated Spanish population as well as in the wild population sampled at the same site (ES21-4 and ES21-5, Supplementary Table S1). The assembled partial genomic scaffold of 2.3kb with an internal gap (estimated at ca. 370 nt) showed 97% nt identity in a BLASTN search with the recently reported parsley umbravirus 1 (PaUV1, OM419177; 3,087nt). However, despite the PaUV1 name, this virus does not show phylogenetic affinities or a genomic organization typical of umbraviruses. In particular, only three ORFs have been predicted for PaUV1, similar to class 2 umbra-like-associated RNAs (Liu et al., 2021) and its phylogenetic clustering, based on a the partial RdRp gene (OM419177, nt 1140-2157) shows that it clusters away from umbraviruses and forms a distinct cluster within the umbra-like-associated RNAs (Figure 2B) suggesting that PaUV1 may represent a separate class of umbra-like-associated RNAs.

Identification of known and novel viruses of the family *Solemoviridae*

Carrot red leaf virus (CtRLV) is a well-known carrot-infecting polerovirus that is part of the complex responsible for the CMD disease affecting carrot crops. CtRLV was detected in most of the sampled carrot populations in France (41/45, 91%) and in all Spanish cultivated populations (5/5, 100%), but not in the wild one. The sequence of a distinct polerovirus, tentatively named wild carrot red leaf virus was deposited in 2020 in GenBank (LT615231). This virus was found exclusively in wild carrot populations, infecting 17/23 of wild populations sampled over two years in France. It was not detected in any of the sampled Spanish populations. A novel enamovirus, tentatively named carrot enamovirus 1 (CaEV-1) and two novel poleroviruses tentatively named carrot polerovirus 1 (CaPV1) and carrot polerovirus 2 (CaPV2) were additionally discovered in various French carrot populations (Table 1).

Following CaPV1 initial identification using the sequencing reads from a wild population sampled in France in 2019 (FR19-15, Supplementary Table S1), plants from that population were individually tested by specific RT-PCR. Three out of the tested 50 plants were positive for CaPV1 and dsRNAs were purified from one of them and sequenced by Illumina. A complete genome could be assembled from this single-plant dataset and the 5' and 3' terminal regions were determined using specific primers in RACE experiments (Supplementary Table S2). The full genome consists of 5,671 nt (average coverage 873x) and shows a genome organization typical for the genus *Polerovirus* with seven predicted ORFs encoding proteins P0 to P5 and P3a (Figure 3). In Blast searches, CaPV1 is most closely related to Trachispermum ammi polerovirus with 89% aa identity in the RdRp and 47-69% to other poleroviruses (Figure 4, Supplementary Table S4). CaPV1 was detected in a high proportion (19/23, 82.6 %) of the French wild carrot populations over the two sampling years irrespective of their geographical origin (Supplementary Table S5). On the other hand, it was not detected in cultivated and off-type French populations, or in the Spanish populations.

A very small contig of ca. 340 nt originally identified from a wild carrot population (FR19-15, Supplementary Table S1) suggested the existence of an additional, distinct polerovirus. In order to obtain a complete genomic sequence, the same strategy as for CaPV1 was followed: plants of the identified population were individually analyzed by a specific RT-PCR-assay (Supplementary Table S2) and the

single positive plant used to extract dsRNAs that were then analyzed by HTS. The sequencing reads thus obtained allowed to assemble a near complete genome of 5.2 kb for CaPV2 (324x average coverage) but no specific efforts were made to determine the precise missing 5' and 3' genome ends. CaPV2 is most closely related to Torrilis crimson leaf virus (GenBank LT615235) and CtRLV (GenBank LC434063) (Figure 4) sharing with them respectively 72% and 71% aa identities in the RdRp (Supplementary Table S4), values that are well under the 10% aa divergence species cut-off threshold. CaPV2 was identified in one cultivated, one off-type and twelve wild French populations but was not identified in Spain. Similar to CaPV1, CaPV2 showed a wide geographical distribution in France, being detected in all five sampled regions (Supplementary Table S5).

Recombination events have been identified between members of the *Polerovirus* genus based on full length genome alignments. Among the viruses identified in the carrot virome, the clearest recombination event concerns wild carrot red leaf virus, which was identified by all algorithms of the RDP4 program. WCtRLV has been identified as recombinant (combined corrected probability $1.2xE^{-113}$) with CtRLV as major parent and CaPV1 as minor parent, the recombined region concerning part of the CP readthrough (CP-RT) domain (Supplementary Figure S1). Another potential recombination event was tentatively identified by six out of seven algorithms for Trachispermum ammi polerovirus, resulting from a recombination with CaPV1 as major parent and Torrilis crimson leaf virus as minor parent. This tentative recombination event was identified in the CP gene, but with a much lower combined corrected probability of $6.9E^{-9}$. A summary of both recombination events is given in Supplementary Figure S1.

A nearly complete genome of a novel enamovirus, CaEV1, of 5.1 kb (average coverage 31x, Table 1) was obtained from a wild carrot population sampled in France in 2019 (FR19_16, Supplementary Table S1). The sequence is most closely related to a partial sequence of Arracacha latent virus E (GenBank MF136435, Figure 4) with respectively 57% aa identity and 56% nt identity in the full RdRp and the nearly full-length genome (Supplementary Table S4). The genome organization of CaEV-1 is typical for the genus, with five predicted ORFs (Figure 3). The first ORF, ORF0, encodes the P0 protein, which has been shown to act as an RNA silencing suppressor (Fusaro et al., 2012). The second ORF, ORF1, harbors a conserved serine-like protease motif (Peptidase S39, Pfam PF02122). The third ORF, ORF2,

is translated by a -1 ribosomal frameshift of ORF1, resulting in an ORF1-ORF2 fusion protein, which encodes the viral RdRp. The fourth and fifth ORFs, ORF3 and ORF4, respectively encode the CP (with a conserved luteovirus CP motif, pfam PF00894) and the CP-RT expressed by readthrough of the ORF3 stop codon, which contains a pfam PF01690 conserved motif and is thought to be involved in aphid transmission (Jolly and Mayo, 1994). CaEV1 lacks a movement protein, which is typical for members of the *Enamovirus* genus. CaEV1 was more prevalent in wild carrot populations, being detected in 11/23 wild populations and in only 2/15 cultivated populations and one off-type population from France and was not detected in Spain. In France, it was geographically limited to southwest France, including Nouvelle Aquitaine and one population sampled in Occitanie (Supplementary Table S5).

Several polerovirus-associated RNAs have been identified in the French and Spanish carrots virome including carrot red leaf virus-associated RNA (CtRLVaRNA), beet western yellows virus-associated RNA (BWYVaRNA), and two other such molecules detected for the first time in carrots, arracacha latent virus E-associated RNA (ALVEaRNA) and a new agent, carrot red leaf virus-associated RNA 2 (CtRLVaRNA-2). These self-replicating subviral agents depend on helper viruses of the family *Solemoviridae* for aphid transmission and have recently been classified as tombus-like associated RNAs (tlaRNAs) due to their phylogenetic affinities to the family *Tombusviridae* (Campbell et al., 2020). A detailed description of their diversity, distribution and preferential association with either wild or cultivated carrots within this metagenomic study has recently been provided (Schönegger et al., 2022a). Moreover, two novel hammerhead ribozyme containing single-stranded circular satellite RNAs carrot red leaf virus satellite 1 (CtRLV Sat1) and carrot red leaf virus satellite 2 (CtRLV Sat2) have also been identified in Spanish and French carrot populations and reported in detail recently (Babalola et al., 2022).

Menzel et al. (2009) had described two linear single-stranded satellite RNAs of respectively 745 and 748 nt and named carrot mottle virus satellite RNA (CMoVsatRNA) and carrot mottle mimic virus satellite RNA (CMoMVsatRNA). These two satellites are closely related, since they show 94% nt sequence identity with each other. Given the diversity of these two satellites and their very close relationship, it was not possible to easily distinguish them and, for the sake of simplicity, all related contigs identified in our study were annotated as representing CMoVsatRNA. It was identified from

most of the French populations (41/44, 93%) and all Spanish populations. In addition, smaller contigs of ~320 nt (average coverage of the retained reference contig 620x) were identified from various French populations and showed a more distant relationship to CMoVsatRNA, with only 84-85% nt identity in BLASTN searches with CMoVsatRNA and CMoMVsatRNA. These contigs are therefore tentatively considered as representing a distinct satellite RNA, for which the name carrot mottle virus satellite RNA 2 (CMoVsatRNA2) is proposed. This additional satellite was preferentially identified in French wild carrot populations (91%, 21/23) as compared to cultivated (27%, 4/15) or off-type (43%, 3/7) populations. It was identified in three regions of France (Supplementary Table S5) but not in Spain.

Known and novel *Betaflexiviridae* members in the carrot virome

Two chordoviruses, carrot chordovirus 1 (CChV1) and carrot chordovirus 2 (CChV2) have been described from carrot samples in the UK (Adams et al., 2014). Both viruses were identified at low frequency in the French carrots virome (CChV1 in one cultivated, one off-type and one wild population; CChV2 in two cultivated and two wild populations).

In addition, two novel chordoviruses, for which the names carrot chordovirus 3 (CChV3) and carrot chordovirus 4 (CChV4) are proposed, have been identified in one off-type population (FR20-8, Supplementary Table S1) and one wild population (FR20-16, Supplementary Table S1), respectively. A 6,860 nt scaffold (39x average coverage) was obtained for CChV3, covering ~84% of the full genome. In GenBank BlastN searches, this contig shows highest homology to hogweed virus 4 (HV4, GenBank OK032418). Pairwise comparisons of CChV3 with HV4 show highest pairwise aa identities of 67% and 69% in the RdRp and CP gene, respectively (Supplementary Table S6, Figures 5A and 5B), which are well below the 80% aa identity species cut-off thresholds for these proteins in the *Betaflexiviridae*. CChV4, identified as a 7.7 kb nearly complete genome with 23x average coverage (Table 1), also shows highest aa pairwise identities to HV4, with respectively 71% and 69% in the RdRp and CP. Similar to lettuce chordovirus 1, the two novel chordoviruses show four predicted ORFs, encoding from 5' to 3' a large, ca. 210 kDa replicase (REP), a 30K-family movement protein (Pfam PF01107), a capsid protein (Pfam PF05892) and a fourth protein (12.3 kDa) of unknown function (Figure 6A). CChV3 and CChV4

share with each other respectively 73% and 69% aa identy in the RdRp and CP and cluster together in the corresponding reconstructed phylogenetic trees (Figures 5A and 5B, Supplementary Table S6).

A tentative novel vitivirus, carrot vitivirus 1 (CaVV1), was identified in a single population of wild carrots sampled in 2020 in Gironde, France (FR20-20, Supplementary Table S1). This is, to our knowledge the first report of a member of the genus *Vitivirus* infecting carrots. A near complete CaVV1 genome scaffold of 7.5 kb with an average coverage of 221x was reconstructed (Table 1). It shares the highest identity with a partial sequence of mint vitivirus 2 (MV-2, GenBank NC_043088) with 83% aa identity (72% nt identity) in the RdRp (Supplementary Table S6). While this value is above the species demarcation criterion, the partial MV-2 RdRp sequence comprises only ~30% of the REP, so that it is not possible to estimate overall REP relatedness, but it is estimated to be lower than the computed 83%, since the available MV-2 sequence covers the most conserved part of the REP protein. The pairwise aa distances in the CP between CaVV1 and MV-2 is 78%, below the species cut-off threshold. The next closest virus in the pairwise comparisons is tomato vitivirus 1 (GenBank OL472239) with respectively 70% and 81% aa identity in the RdRp and CP (Supplementary Table S6). CaVV1 shows a typical vitivirus genome organization including five ORFs encoding from 5' to 3' a replicase, a 20K-protein of unknown function typical of genus members, a movement protein, a capsid protein and a nucleotide binding protein (Figure 6B). Interestingly, CaVV1 ORF2 does not have a canonical AUG codon but could be expressed from an alternative CUG start codon.

Known and novel closteroviruses in the carrot virome.

Carrot yellow leaf virus (CYLV), a known member of the *Closterovirus* genus, was identified in one cultivated population sampled in 2019 in Gironde (FR19-10, Supplementary Table S1) and in one wild population from north of France (FR20-22, Supplementary Table S1). Carrot closterovirus 1 (CtCV1), previously described in the UK (Adams et al., 2014), was also identified in the same wild carrot population from north of France. Two additional novel viruses of the *Closterovirus* genus were identified from a few French or Spanish carrot populations. A large contig covering a nearly complete genome (16.3 kb, average coverage 24,536x, Table 1) was obtained from a cultivated carrot population (FR20-9, Supplementary Table S1) and corresponds to a novel closterovirus for which the name carrot

closterovirus 2 (CtCV2) is proposed. A second large contig of 14.6 kb was initially assembled from a different cultivated population (FR20-11, Supplementary Table S1) sampled in 2020 in France. The contig was extended by repeated mapping of residual reads to obtain a near complete genome scaffold of 16.3 kb (4,173x average coverage, Table 1) for a second novel closterovirus, for which the name carrot closterovirus 3 (CtCV3) is proposed. CtCV3 was identified in two French populations, one cultivated from Nouvelle Aquitaine and one wild from Ile-de-France (20-11 and 20-16, Supplementary Table S1), and in four cultivated Spanish populations (ES21-1B, ES21-2, ES21-3 and ES21-5, Supplementary Table S1). These two novel closteroviruses show a genomic organization with 10 predicted ORFs typical for the genus. The 5' ORFs 1a and 1b encode replication associated proteins: P1a contains RNA methyltransferase (MET) and RNA helicase (HEL) domains, while P1b, expressed by a +1 ribosomal frameshift of ORF1a encodes RdRp. The 3' proximal ORFs are expressed from monocistronic subgenomic (sg) RNAs with five downstream genes that are highly conserved among *Closteroviridae* members (Dolja et al., 2006) including a ~6 kDa hydrophobic protein (p6), a HSP70 homolog (HSP70h), a ~60 kDa protein (p60); the minor, duplicate capsid protein (CPd), and the major capsid protein (CP). The two novel closteroviruses, CtCV2 and CtCV3, encode an additional ~30 kDa protein of unknown function which is also present in the CYLV genome and in a few other closteroviruses, such as citrus tristeza virus (CTV, GenBank U16304) or beet yellow stunt virus (BYSV, GenBank U51931) (Figure 7). This gene is, however, apparently absent from the genome of CtCV1, which presents a more typical genomic organization. A unique feature of CYLV, CtCV2 and CtCV3 is the location of the p30 ORF, which is found dowstream the p6 ORF protein, whereas in CTV and BYSV, it is found directly downstream ORF1b (Figure 7). In addition, a unique genomic feature of the two novel closteroviruses is the position of the HSP70h ORF, which is located directly downstream of ORF1b, when in all other closteroviruses, it is located downstream the p6 ORF (Figure 7). Lastly, the genomes of the two novel closteroviruses contain two 3' terminal ORFs encoding respectively p21 and p20 proteins whose homologs in BYSV have been shown to be involved in systemic movement and suppression of RNA silencing, respectively (Dolja et al., 2006).

The two novel closteroviruses share respectively 83% and 61% aa identity in their RdRp and in the full ORF1a-ORF1b fusion protein, respectively. Pairwise comparisons in the HSP70h and the CP show respectively 80% and 70% aa identity between CtCV2 and CtCV3 (Supplementary Table S7). Both viruses share 74-78% aa identity, 65-68% HSP70h aa identity and 44-47% CP aa identity CtCV1 and CYLV. Even though the pairwise distances in the RdRp and HSP70h gene show that the two novel closteroviruses are related, their distance is above the ICTV species demarcation threshold of 25% aa divergence in the CP. Altogether, these comparisons suggest that the assembled genomes correspond to viruses that belong to two distinct and novel species.

A new alphaendornavirus in a wild carrot population

A novel alphaendornavirus was identified from a wild carrot population sampled in the north of France in 2020 (FR20-18, Supplementary Table S1). According to our knowledge, this is the first record of a virus of the *Endornaviridae* family infecting carrots. The assembled genomic scaffold of 14.6 kb (average coverage 608x, Table 1) represents a near complete genome and shows highest phylogenetic affinities to winged bean alphaendornavirus 1 (WBEV1, GenBank NC_031336) and fagopyrum esculentum endornavirus 1 (FeEV1, GenBank LC500285) with 54-55 % pairwise nt identity and 44-45% pairwise aa identity over the genome polyprotein (Supplementary Table S8). CaAEV1 clusters together with these two viruses and with related members of the genus *Alphaendornavirus*, separately from members of the *Betaendornavirus* genus but its divergence level with WBEV1 and FeEV1 unmistakeably make it a novel species (Figure 8).

Novel *Spinareoviridae* identified in wild carrot populations in France

Two novel viruses, carrot reovirus 1 (CaRV1) and carrot reovirus 2 (CaRV2), have been identified in wild carrot populations in France. Contigs for ten RNA genomic segments could be obtained for CaRV1, which was identified in three wild populations (FR20-18, FR20-20, FR20-21, Supplementary Table S1) sampled in 2020 in France. On the other hand, only three genomic segments, including one encoding the viral RdRp, could be identified by Blast searches for CaRV2, which was identified in only two populations sampled in different parts of the north of France (populations FR20-19 and FR20-22, Supplementary Tables S1 and S5). Both viruses show homologies to members of the family

Spinareoviridae (previously *Reoviridae*) and group in separate subclusters of a clade composed of rice ragged stunt virus (RRSV, *Oryzavirus*) and unclassified plant-associated viruses, including grapevine Cabernet Sauvignon reovirus (GCSV) and raspberry latent virus (RpLV) (Figure 9). The latter was proposed to belong to a novel genus in the previously listed *Reoviridae* family (Quito-Avila et al., 2011). Eight out of 10 obtained CaRV1 genomic segments appear to be full length and show nearly identical conserved ends as seen for RpLV (Quito-Avila et al., 2011) with the conserved tetranucleotide AGUU at the 5' termini and the hexanucleotide GAAUAC at the 3' termini. Pairwise RdRp aa identities of respectively 71% and 66% between CaRV1 and GCSV or RpLV, distances that are well below the 86-94% aa identities registered between maize rough dwarf virus, southern rice black-streaked dwarf virus and rice black streaked dwarf virus, which are all ICTV accepted species (Supplementary Table S9). If a new genus is created at some point to include RpLV, CaRV1 would clearly be a candidate to be integrated in this genus, together with GCSV (Figure 9). In the case of the partial CaRV2 RdRp, the highest aa identity recorded, 68%, was with Hubei reo-like virus 5 (KX884720), a virus identified from insect metagenomic data.

Other double-stranded RNA viruses identified in the carrot virome

A few viruses of the family *Partitiviridae* have been reported to infect carrots. A member of the genus *Alphapartitivirus*, carrot cryptic virus (CaCV), has been reported in Germany (Willenborg et al., 2009) and unassigned *Partitiviridae* members have been reported in Japan (Natsuaki et al., 1990), referred to as carrot temperate viruses 1-4 (CteV1-4). Unfortunately, no reference isolates or sequence information are available for the latter. In the 9th Report of Virus Taxonomy (Ghabrial et al., 2012), CteV1, CteV2 and CteV4 have been listed as members of the *Alphapartitivirus* genus and CteV3 in the genus *Betapartitivirus*. CaCV has been detected in a high proportion of cultivated (11/15, 73.3%) and off-type carrot populations (6/7, 86%) but was only detected in one wild population sampled near Bordeaux (FR20-15, Supplementary Table S1). Three other viruses from different *Partitiviridae* genera were also identified. A divergent variant of dill cryptic virus 2 (DiCV2) was detected in all three types of carrot populations, infecting a higher proportion of wild carrot populations (6/23, 39%) than cultivated (2/15, 8%) or off type ones (1/7, 14%). It was geographically limited to the region Nouvelle-Aquitaine

(Supplementary Table S5). The DiCV2 carrot isolate here identified shows significant divergence from the type DiCV2 from dill in GenBank (NC_021147 and NC_021148) showing with it only 88% and 86% pairwise aa identity in the RdRp and CP, respectively (Supplementary Table S10), which are borderline values for species demarcation according to the ICTV guidelines for the genus *Betapartitivirus* (Ghabrial et al., 2012). Phylogenetically, DiCV2 isolates cluster together with other betapartitiviruses and are clearly separated from other genera in the *Partitiviridae* family (Figure 10A). Two novel members of the *Deltapartitivirus* genus have also been identified in the carrot virome, representing to our knowledge the first report of deltapartitiviruses in carrot. These two novel viruses are referred to as carrot cryptic virus 2 (CaCV2) and carrot cryptic virus 3 (CaCV3) and typically possess two RNA segments of about 1.5 kb each. RNA1 codes for the viral RdRp and RNA2 for the CP. CaCV2 was detected in 8/23 (34.8%) cultivated and 4/15 wild populations (26.7%), both in north and southwest France (Supplementary Table S5). It was also detected in one cultivated Spanish population (ES21-1A). The two contigs assembled for the genomic RNAs of CaCV2 have an average coverage of 31,529x (Table 1). CaCV2 is most closely related to persimmon cryptic virus (GenBank HE805113-14) and, based on the RdRp, clusters with other members of the *Deltapartitivirus* genus (Figure 10A). A similar clustering was observed when using the CP (not shown). The second novel deltapartitivirus CaCV3 was identified from a single wild carrot population in Spain (ES21-4, Supplementary Table S1). Similarly, to CaCV2, the genomic contigs assembled had very deep average coverage (13,634x, Table 1) and CaCV3 shows clear phylogenetic affinities to members of the genus *Deltapartitivirus* (Figure 10A). Highest pairwise RdRp aa identity (68%) was observed with persimmon cryptic virus (YP_006390090), while the corresponding value for the less conserved CP was only of 42% with rhodiola cryptic virus 1 (QED42889) (Supplementary Table S10). However, these two novel viruses cluster in different subgroups within the *Deltapartitivirus* clade and their RdRp and CP proteins are very divergent (Figure 10A, Supplementary Table S10). Given the very limited information available on CTeV1-4, it is not possible to ascertain whether they might be similar to CaCV2 and CaCV3 reported here.

Contigs showing similarities to members of the genus *Amalgavirus* (family *Amalgaviridae*) were detected in several French wild and off-type populations in both sampling years. The virus was not identified from French cultivated populations or from carrots sampled in Spain. A nearly complete

genome of this new amalgavirus, referred to as carrot amalgavirus 1 (CaAV1), was obtained from a wild carrot population sampled in 2020 in France (FR20-15, Supplementary Table S1). The obtained 3.2 kb contig (average coverage 929x, Table 1) harbors the two ORFs characteristic of amalgaviruses. ORF1 encodes a 44 kDa protein (398 aa) of unknown function and ORF2, translated in the +1 frame with respect to ORF1, codes for a 118 kDa (1066 aa) fusion protein containing the RdRp. The overlapping ORF1-ORF2 sequence of CaAV1 contains the +1 ribosomal frameshift motif conserved among amalgaviruses (UUU CGN) (Nibert et al., 2016). CaAV1 does not seem to be closely related to any previously reported amalgavirus and shows highest RdRp pairwise aa identity of 48% with Gevuina avellana amalgavirus 1 (DAB41733, Supplementary Table S11).

A short contig of 777 nt (19x average coverage, Table 1) showing sequence similarity to members of the genus *Amalgavirus*, was assembled from the reads from the wild population sampled in Spain (ES21-4, Supplementary Table S1). The sequence covers a central region of amalgavirus ORF2 and is most closely related to CaAV1 with 69% aa identity (Supplementary Table S11), which is in the range for species demarcation (65-70% aa identity in the RdRp) for the genus *Amalgavirus*. This contig is therefore considered as representing a partial genome of a distinct amalgavirus, with the proposed name carrot amalgavirus 2 (CaAV2). Even though the sequence of CaAV2 is only partial, raising uncertainties about the divergence level when analyzing complete proteins, pairwise identity levels between recognized amalgavirus species in the region encompassed by the available CaAV2 sequence are similar to the CaAV1-CaAV2 divergence level, providing evidence that the two sequences correspond to different species (data not shown). Both viruses cluster unambiguously within the *Amalgavirus* genus and away from zygosaccharomyces bailii virus Z (GenBank NP_624325), the only member of the second genus recognized in the *Amalgaviridae* family (genus *Zybavirus*) (Figure 10B).

Two overlapping contigs of 923 nt and 4,254 nt were assembled from the reads from an off-type carrot population sampled in 2020 in Gironde (20-4, Supplementary Table S1). The contigs were scaffolded into a near complete genome of 5,128 nt (8876x average coverage, Table 1). This contig harbors two ORFs: ORF1 encodes a 793 aa protein corresponding to the CP, and ORF2, which overlaps ORF1 by only one nucleotide, encodes a 69 kDa (855aa) protein with a RdRp conserved motif. This genomic

organization is similar to that of viruses of the *Totiviridae* family, and the encoded proteins show affinities to those of *Toviviridae* members, so that it is proposed to refer to the identified virus as carrot-associated toti-like virus (CaaTLV). In a BLASTN search, CaaTLV shares highest identities with black raspberry F virus (BRVF, GenBank EU082131). Pairwise comparisons of its CP and RdRp confirm BRVF as the most closely related virus with 65% and 73% aa identity in the CP and RdRp, respectively (Supplementary Table S12) and phylogenetic analyses demonstrate its affinities with a group of plant-associated totiviruses (Figure 10C). Although the pairwise protein identity levels are above the accepted molecular demarcation criteria (50% aa identity threshold), the viruses in question originate from different host species, which is a biological ICTV demarcation criterion. In addition, RdRp identity levels between accepted plant-associated totiviruses are often well above the 50% threshold (e.g. between peach-associated virus 2 (QSV39138) and loquat-associated totivirus 1 (OK318989) with 71% aa identity, Supplementary Table S12), all providing evidence that CaaTLV may constitute a new species within the genus *Totivirus*.

Unclassified putative novel *Riboviria* in the carrot virome

Several large ca. 8 kb nearly identical (ca. 99-100% nt identity) contigs were obtained from different French cultivated and off-type carrot populations. These contigs span the nearly complete genome of a novel virus showing distant homologies to members of the *Benyviridae* family. The selected reference contig 8,068 nt long (average coverage 2927x, Table 1) harbours two predicted ORFs, the largest encoding a 244 kDa protein (2,201 aa) with viral RNA helicase (PF01443) and RdRp (PF00978) motifs and the smallest, 3' located, encoding a 18.5 kDa protein (165 aa) of unknown function (Figure 11A). The genome 3’end is polyadenylated, but the length of the polyA sequence was not determined. This genome organization is similar to that several other unclassified viruses in GenBank with homology to *Benyviridae* members (Figure 11B). In a BLASTN search, two recently released GenBank 1.2-1.3kb contigs (OM419188 and OM419189), referred to as carrot-associated RNA virus 1 (CaRNAV1) and obtained from a historical carrot sample collected in 2005 in Australia, show 99% and 100% identity to the sequence of the virus identified here, indicating that it should be referred to as carrot-associated RNA virus 1. The next most closely related viruses are red clover virus 1 (GenBank MG596242) and

Dactylorhiza hatagirea beny-like virus (BK013327), which show pairwise RdRp aa identity levels of 43% and 39%, respectively (Supplementary Table S13). Accordingly, CaRNAV1 groups together red clover RNA virus 1 and other unclassified viruses in a separate cluster from accepted *Benyviridae* members and shows more distant relationships to other rod shaped viruses of the *Virgaviridae* family (Figure 11B). The virus was very frequently detected in the collected field carrot populations in both countries, infecting 100% of cultivated and off-type populations in 2019 and 82% of cultivated and 100% of off type populations sampled in 2020 in France as well as all four cultivated Spanish carrot populations.

Further identified putative novel Riboviria comprise carrot flavi-like virus 1 (CtFLV1) and carrot Ker-like virus (CaKLV). The molecular characterization of CtFLV1 and its geographical distribution has recently been described by Schönegger et al. (2022b), while CaKLV, which shows homologies to viruses first identified in a viral metagenome from the Kerguelen island and together likely represent a potential higher order viral taxon will be presented elsewhere (TC, AM, DS et al., manuscript in preparation).

DISCUSSION

In the past decade, several studies have begun to explore the largely underestimated viral diversity in unmanaged ecosystems and the role of wild plants as virus reservoirs (Muthukumar et al., 2009; Roossinck, 2012; Thapa et al., 2015; Bernardo et al., 2018; Susi et al., 2019; Ma et al., 2020; Ma et al., 2021; Maachi et al., 2022; Rivarez et al., 2022). The carrot pathosystem, consisting of cultivated and wild carrot populations, is particularly interesting with respect to metagenomic comparisons of virus communities due to the low genetic barrier to virus flow since both carrot types belong to the same species. This situation is further enriched by off-type carrot populations that most likely represent hybrids between cultivated and wild carrots contaminating commercial carrot seed lots (Magnussen and Hauser, 2007). In a large-scale effort, we characterized the virome of 15 cultivated, 23 wild and seven off-type carrot populations from five different regions in France over two years. In addition, five cultivated and one wild carrot populations were sampled in Spain. Using a highly purified dsRNA-based HTS approach, we identified a very rich virome including 45 viruses of which 25 are novel (Figure 1C).

In the majority of cases, very long contigs, often representing near full length genomes could be assembled from the reads derived from the pools of plants used, thanks to the deep average genome coverage achieved. With four exceptions, a coverage in excess of 150x was achieved for all novel viruses reported here (Table 1). This depth of coverage and ability to assemble very long contigs parallel results obtained in other comparable virome studies (Ma et al., 2020; Rivarez et al., 2022).

In a limited screening of public carrot RNASeq transcriptomic data in GenBank sequence reads archives (SRAs), a total of seven viruses were identified, five of which are in common with the present study, with the other two identified viruses being the potyvirus carrot thin leaf virus and a novel *Partitiviridae* member distinct from those reported here (data not shown). The range of carrot infecting viruses with dsRNA genomes was extended through the identification of novel members of the families *Amalgaviridae, Totiviridae*, *Spinareoviridae* as well as two novel viruses of the *Deltaparitivirus* genus and one new host record for a divergent isolate of the betapartitivirus DiCV2. Most of the novel viruses were +ssRNA viruses that are aphid transmitted (Figure 1C) including members of the *Closteroviridae*, *Solemoviridae* and *Tombusviridae* families, as well as their associated subviral agents, that are known to be involved in complex systems of mutual complementation and assistance. Umbraviruses (family *Tombusviridae*), that lack a CP gene, do not form virus particles and therefore rely on helper viruses, characteristically from the family *Solemoviridae* (former *Luteoviridae*), for encapsidation and aphid transmission (Ryabov et al., 2012). The carrot motley dwarf disease complex is based on such an interaction between a polerovirus, CtRLV, and umbraviruses (CMoV or CMoMV), often in further association with an associated RNA (CtRLVaRNA). In this study, we identified a very diverse and rich spectrum of poleroviruses, umbraviruses and associated subviral agents. A total of four different poleroviruses was identified, two known (CtRLV, WtRLV) and two novel (CaPV1 and CaPV2), plus a new enamovirus (CaEV1). In addition, six different umbraviruses, three known (CMoV, CMoMV, WCMoV) and three new or potentially new (CaUV2 and CaUV3, CaUV1) were detected in different carrot populations in France, often in complex co-infections, when single plants were analyzed. The carrot virome was equally rich in subviral agents, with four different polerovirus-associated RNAs (also referred to as tombus-like-associated RNAs, tlaRNAs) identified in French and Spanish carrot

populations (Schönegger et al., 2022a). In addition, both linear and small circular, hammerhead-containing satellite RNAs (Babalola et al., 2022) were detected. Synergistic interaction between poleroviruses and umbraviruses have been reported with destructive effects on different crops (Naidu et al., 1998, Mo et al., 2002) but there is little information on their precise interactions and even less so on the level of specificity or promiscuousness involved in these interactions. Recent studies investigating disease complexes under field condition parallel our findings of novel polero- and umbraviruses in coinfections with known viruses that are involved in helper-dependent disease complexes such as the CMD complex. Abraham et al. (2014) showed that a novel umbravirus and a novel satellite in mixed infections with the polerovirus potato leafroll virus (PLRV), were causing the Tobacco Bushy Top Disease (TBTD) in Ethiopia. Moveover, two novel poleroviruses, tobacco polerovirus 1 and tobacco polerovirus 2, have been identified in symptomatic TBTD affected plants coinfected with other viruses such as tobacco bushy top virus (genus *Umbravirus*), tobacco distorting vein virus (genus *Polerovirus*) and their associated satellites in China's Yunnan province (Tan et al., 2021). To gain deeper insights into virus prevalence and co-infection patterns, individual carrot plants of one cultivated and one wild population sampled in France in 2019 (FR19-7 and FR19-9, Supplementary Table S1) were individually tested by specific RT-PCR and analysed by dsRNA-based HTS (DS, TC, AM et al., in preparation). The results confirmed that plants were frequently coinfected by different known and novel poleroviruses, umbraviruses and associated subviral agents, suggesting complex and flexible interactions. Indeed, while arracacha latent virus E (ALVE, genus *Enamovirus*) was not observed in French or Spanish carrot viromes, arracacha latent virus E-associated RNA (ALVEaRNA) was frequently identified, often in coinfection with CtRLV which likely fulfils a helper virus role, thus indicating some degree of plasticity in associations between different helper and dependent viruses (Schönegger et al., 2022a). Co-infections likely explain that several tentative recombination events could be detected among members of the *Polerovirus* genus, including a statistically highly significant recombination in the CP-readthrough (CP-RT) protein involving WCtRLV, CtRLV and the novel CaPV1. The CP-RT domain of the polerovirus genome has been shown to be the most variable genome part, with a high frequency of sites subject to positive selection, and has been suggested to be a major player in mediating adaptation to host plants

and vectors (Latourrette et al., 2021). Indeed, WCtRLV, together with CaPV1, was found exclusively in wild carrot populations, which in turn may be indicative of host adaptation.

Two closteroviruses were previously known from carrot, CYLV and CtCV1. CYLV was first isolated from carrots showing leaf yellowing in Japan (Yamashita et al., 1976) and is semi-persistently transmitted by *Cavariella* spp. (Menzel et al., 2009). CYLV has 10 ORFs similar to beet yellows virus (BYV), but in comparison to other closteroviruses, the genome positions of the p6 and p30 ORFs are reversed. The second closterovirus, CtCV1 was identified in the frame of a diagnostic approach investigating internal root necrosis in carrot crops in the UK using HTS (Adams et al., 2014). In comparison to CYLV, CtCV1 lacks the p30 gene (Figure 7). We have identified two additional, closely related closteroviruses, CtCV2 and CtCV3 with an original genomic organization characterized by an HSP70h ORF located directly downstream of ORF1b (Figure 7). As pointed out by Dolja et al. (2006), five genes downstream the RdRp gene are highly conserved among members of the *Closteroviridae* family (p6, HSP70h, p60, CP and CPm) but their order is generally not conserved. A similar gene reshuffling is seen for example in the genera *Ampelovirus* and *Crinivirus* with a reversed order of the CPd and CP ORFs (Dolja et al., 2006).

Another group of +ssRNA viruses identified is this study comprises the family *Betaflexiviridae*. The genus *Chordovirus* was recently accepted within the family *Betaflexiviridae* (https://ictv.global/taxonomy) and comprises carrot chordovirus 1 and 2 (CChV1 and CChV2), identified in the UK (Adams et al., 2014), lettuce chordovirus 1 identified in lettuce in France (Svanella-Dumas et al., 2018) and hogweed virus 4, a sequence recently submitted in GenBank (OK032418) and obtained from a historical UK hogweed sample. We identified two novel chordoviruses, CChV3 and CChV4 as well as a novel vitivirus, CaVV1. Many viruses of the *Betaflexiviridae* family have no known insect or fungal vector. However, carlaviruses are generally transmitted by aphid vectors, while vitiviruses have been reported to be transmitted by mealybugs (Herrbach et al., 2017) and by the aphids *Ovatus crataegarius* and *Cavariella* spp. (Tzanetakis et al., 2007). Multiple studies have suggested that presence of a closterovirus or ampelovirus is required for insect-mediated transmission of vitiviruses (Tzanetakis et al., 2007; Bem and Murant, 1979), for a review see Maree et al., 2020. No infection by

closteroviruses was however detected in the carrot population in which CaVV1 was identified and it will be therefore of much interest to further investigate the transmission mode of the identified *Betaflexiviridae* family members in carrot populations and their potential dependence on other viruses for transmission.

Previously unknown putative novel Riboviria have been found in different carrot populations including the recently described CtFLV1 (Schönegger et al., 2022b), carrot-associated RNA virus 1 and carrot Ker-like virus (TC, AM, DS et al., in preparation). Two ca. 1.2 kb contigs of carrot-associated RNA virus 1 have been identified in a 17 years old carrot sample from Australia (Fox et al., 2022) that are nearly identical to the near complete genomes we assembled for this virus exclusively identified in cultivated and off-type carrot populations from France and Spain. This suggests a very limited diversity of this virus in both space and time. CaRNAV1 was also identified from two public SRA datasets, SRR5829255 (cultivated carrot cv. Kurada sequenced in China) and SRR11243946 (cultivated carrot sequenced in Sweeden) (data not shown). A RT-PCR screening of seedlings germinated from carrot seeds from different origins has confirmed this low diversity and shown CaRNAV1 to be seed-transmissible in carrot (DS, TC, AM et al., in preparation).

Most of the identified novel viruses showed preferential associations with wild carrots, either occurring exclusively in wild populations or infecting only a small proportion of cultivated populations (Table 1). This suggests that wild plants may represent a reservoir of viral diversity from which novel viruses attacking carrot crops could potentially emerge, but also suggests that despite the close taxonomic relationship linking cultivated and wild carrots, host adaptation or other unforeseen barriers may limit flow of viruses between plants belonging to these two subspecies. Using HTS we have obtained a comprehensive picture of the carrot virome at the agroecological interface in two countries, France and Spain, which proved to be very diverse and rich, especially in viruses involved in complex interactions of mutual interdependencies for aphid transmission. Our results show that the carrot pathosystem provides a particularly interesting playground for further investigations of the poorly studied interplay between coinfecting helper and dependent viruses and of the evolutionary processes leading to synergistic or antagonistic virus-vector-host relationships under natural field conditions.

Acknowledgements: The authors wish to thank the INRAE GetPlaGe platform (GenoToul Toulouse, France) for Illumina sequencing, and S. Plas (Invenio, France), S. Dandin (Ctifl Sileban, France), and F. Villeneuve (Ctifl Lanxade, France) for arranging access to carrot crops. We also E. Geoffriau and the Carrot and other Apiaceae Biological Resources Center (IRHS, INRAE-Institut Agro-Université Angers, Angers, France) for valuable exchanges and for seeds of different carrot genotypes. We thank UK collegues Neil Boonham, Adrian Fox and Ferran Pamblanco Salavert at the University of Newcastle and FERA for their valuable inputs and exchanges on identified carrot viruses in the UK.

Funding: This study was funded by the European Union through a Horizon 2020 Marie Skłodowska-Curie Actions Innovative Training Network (H2020 MSCA- 60 ITN) project "INEXTVIR" (GA 813542).

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

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

REFERENCES

Abraham AD, Menzel W, Bekele B, Winter S (2014) A novel combination of a new umbravirus, a new satellite RNA and potato leafroll virus causes tobacco bushy top disease in Ethiopia. *Arch. Virol.* **159**, 3395-3399. https://doi.org/10.1007/s00705-014-2202-4

Adams IP, Skelton A, Macarthur R, Hodges T, Hinds H, Flint L, Nath PD, Boonham N, Fox A (2014) Carrot yellow leaf virus is associated with carrot internal necrosis; *PLoS ONE*, **9**, e109125. https://doi.org/10.1371/journal.pone.0109125

Alexander HM, Mauck KE, Whitfield AE, Garrett KA, Malmstrom CM (2014) Plant-virus interactions and the agro-ecological interface. *Eur. J. Plant Pathol.* **138**, 529-547. https://doi.org/10.1007/s10658-013-0317-1

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990) Basic local alignment search tool. *J Mol Biol.* **215**, 403-410. https://doi: 10.1016/S0022-2836(05)80360-2

Anderson PK, Cunningham AA, Patel NG, Morales FJ, Epstein PR, Daszak P (2004) Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends Ecol. Evol.* **19**, 535-544. https://doi.org/10.1016/j.tree.2004.07.021

Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, De Castro E, Duvaud S, Flegel V, Fortier, A, Gasteiger E (2012) ExPASy: SIB bioinformatics resource portal. *Nucl. Acids Res.* **40**, W597–W603. https://doi.org/10.1093/nar/gks400

Babalola BM, Schönegger D, Faure C, Marais A, Fraile A, Garcia-Arenal F, Candresse T (2022) Identification of two novel putative satellite RNAs with hammerhead structures in the virome of French and Spanish carrot samples. *Arch. Virol.* **167**, 2287-2292. https://doi.org/10.1007/s00705-022-05538-z

Bem F, Murant AF (1979) Transmission and differentiation of six viruses infecting hogweed (*Heracleum sphondylium*) in Scotland. *Ann. Appl. Biol.* **92**, 237-242. https://doi.org/10.1111/j.1744-7348.1979.tb03869.x

Bernardo P, Charles-Dominique T, Barakat M, Ortet P, Fernandez E, Filloux D, Hartnady P, Rebelo TA, Cousins SR, Mesleard F, Cohez D, Yavercovski N, Varsani A, Harkins GW, Peterschmitt M, Malmstrom CM, Martin DP, Roumagnac P (2018) Geometagenomics illuminates the impact of agriculture on the distribution and prevalence of plant viruses at the ecosystem scale. *ISME Journal* **12**, 173-184. https://doi.org/10.1038/ismej.2017.155

Brunt AA, Crabtree K, Dallwitz MJ, Gibbs AJ, Watson L, Zurcher EJ (eds.) (1996). Plant Viruses Online: Descriptions and Lists from the VIDE Database. Version: 20th August 1996. http://biology.anu.edu.au/Groups/MES/vide/

Campbell AJ, Erickson A, Pellerin E, Salem N, Mo X, Falk BW, Ferriol I (2020) Phylogenetic classification of a group of self-replicating RNAs that are common in co-infections with poleroviruses. *Virus Res.* **276**, 197831. https://doi.org/10.1016/j.virusres.2019.197831

Dolja VV, Kreuze JF, Valkonen JPT (2006) Comparative and functional genomics of closteroviruses. *Virus Res.* **117**, 38-51. https://doi.org/10.1016/j.virusres.2006.02.002

Elena SF, Fraile A, García-Arenal F (2014) Evolution and Emergence of Plant Viruses. *Adv. Virus Res.* **88**, 161-191. https://doi.org/10.1016/B978-0-12-800098-4.00003-9

Fox A, Gibbs AJ, Fowkes AR, Pufal H, McGreig S, Jones RAC, Boonham N, Adams IP (2022) Enhanced Apiaceous Potyvirus Phylogeny, Novel Viruses, and New Country and Host Records from Sequencing *Apiaceae* Samples. *Plants* **27,** 1951. https://doi: 10.3390/plants11151951

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. In: Viral Metagenomics: Methods and Protocols, Pantaleo V & Chiumenti M (Eds.), Humana Press, New York (USA), pp 77-95.

Fusaro AF, Correa RL, Nakasugi K, Jackson C, Kawchuk L, Vaslin MFS, Waterhouse PM (2012) The Enamovirus P0 protein is a silencing suppressor which inhibits local and systemic RNA silencing through AGO1 degradation. *Virology* **426**, 178-187. https://doi.org/10.1016/j.virol.2012.01.026

Ghabrial SA, Nibert ML, Maiss E, Lesker T, Baker TS, Tao YJ (2012) Family *Partitiviridae*. In: Virus Taxonomy - Ninth Rep. Int. Committee Taxonomy of Viruses. King AMQ, Adams, MJ, Carstens EB, Lefkowitz EJ, Eds. Elsevier Academic Press, London (UK), pp 523-534.

González R, Butković A, Escaray FJ, Martínez-Latorre J, Melero Í, Pérez-Parets E, Gómez-Cadenas A, Carrasco P, Elena SF (2021) Plant virus evolution under strong drought conditions results in a transition from parasitism to mutualism. *Proc. Natl. Acad. Sci. USA* **118**, e2020990118. https://doi.org/10.1073/pnas.2020990118

Herrbach E, Alliaume A, Prator CA, Daane KM, Cooper ML, Almeida RPP (2017) Vector transmission of grapevine leafroll-associated viruses. In: Grapevine viruses: molecular biology, diagnostics and management, Meng B, Martelli GP, Golino DA, Fuchs M (eds), Springer, Cham 'Switzerland), pp 483-503.

Jolly CA, Mayo MA (1994) Changes in the amino acid sequence of the coat protein readthrough domain of potato leafroll luteovirus affect the formation of an epitope and aphid transmission. *Virology* **201**, 182-185. https://doi.org/10.1006/viro.1994.1283

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol*. **35,** 1547-1549. https://doi: 10.1093/molbev/msy096

Latourrette K, Holste NM, Garcia-Ruiz H (2021) Polerovirus genomic variation. *Virus Evol.* **7**, 1-18. https://doi.org/10.1093/ve/veab102

Lefeuvre P, Martin DP, Elena SF, Shepherd DN, Roumagnac P, Varsani A (2019) Evolution and ecology of plant viruses. *Nat. Rev. Microbiol.* **17**, 632-644. https://doi.org/10.1038/s41579-019-0232-3

Liu J, Carino E, Bera S, Gao F, May JP, Simon AE (2021) Structural analysis and whole genome mapping of a new type of plant virus subviral RNA: Umbravirus-like associated RNAs. *Viruses* **13**, 646. https://doi.org/10.3390/v13040646

Ma Y, Marais A, Lefebvre M, Faure C, Candresse T (2020) Metagenomic analysis of virome cross-talk between cultivated *Solanum lycopersicum* and wild *Solanum nigrum*. *Virology* **540**, 38-44. https://doi:.org/10.1016/j.virol.2019.11.009

Ma Y, Tania Fort T, Marais A, Lefebvre M, Sébastien Theil S, Vacher C, Candresse T (2021) Leaf-associated fungal and viral communities of wild plant populations differ between cultivated and natural ecosystems. *Plant Environ. Interact.* **2**, 87-99. https://doi.orgh/10.1002/pei3.10043

Maachi A, Donaire L, Hernando Y, Aranda MA (2022) Genetic Differentiation and Migration Fluxes of Viruses from Melon Crops and Crop Edge Weeds. *J. Virol.* **96**, e0042122. https://doi.org/10.1128/jvi.00421-22

Magnussen LS, Hauser TP (2007) Hybrids between cultivated and wild carrots in natural populations in Denmark. *Heredity* **99**, 185–192. https://doi.org/10.1038/sj.hdy.6800982

Malmstrom CM, Alexander HM (2016) Effects of crop viruses on wild plants. *Curr. Opin. Virol*. **19**, 30-36. https://doi.org/10.1016/j.coviro.2016.06.008

Marais A, Svanella-Dumas L, Barone M, Gentit P, Faure C, Charlot G, Ragozzino A, Candresse T (2011) Development of a polyvalent RT-PCR detection assay covering the genetic diversity of Cherry capillovirus A (CVA). *Plant Pathol.* **61**, 195-204. https://doi.org/10.1111/j.1365-3059.2011.02488.x

Marais A, Faure C, Bergey B, Candresse T (2018) Viral double-stranded RNAs (dsRNAs) from Plants: Alternative nucleic acid substrates for high-throughput sequencing. In: Viral Metagenomics: Methods in Molecular Biology, Pantaleo V, Chiumenti M (Eds.), Humana Press, New York (USA), pp 45-53.

Maree HJ, Blouin AG, Diaz-Lara A, Mostert I, Al Rwahnih M, Candresse T (2020) Status of the current vitivirus taxonomy. *Arch. Virol*. **165**, 451-458. https://doi.org/10.1007/s00705-019-04500-w

Martin DP, Murrell B, Golden M, Khoosal A, Muhire B (2015) RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evol.* **1**, vev003. https://doi.org/10.1093/ve/vev003

Menzel W, Goetz R, Lesemann DE, Vetten HJ (2009) Molecular characterization of a closterovirus from carrot and its identification as a German isolate of carrot yellow leaf virus. *Arch. Virol.* **154**, 1343-1347. https://doi.org/10.1007/s00705-009-0428-3

Menzel W, Vetten HJ (2008) Complete nucleotide sequence of an isolate of the *Anthriscus* strain of *Parsnip yellow fleck virus*. *Arch. Virol.* **153**, 2173-2175. https://doi.org/10.1007/s00705-008-0240-5

Mihara T, Nishimura Y, Shimizu Y, Nishiyama H, Yoshikawa G, Uehara H, Hingamp P, Goto S, Ogata H (2016) Linking virus genomes with host taxonomy. *Viruses* **8**, 66. https://doi.org/10.3390/v8030066

Mo XH, Qin XY, Tan ZX, Li TF, Wu JY, Chen HR (2002) First report of tobacco bushy top disease in China. *Plant Dis.* **86**, 74. https://doi.org/10.1094/PDIS.2002.86.1.74B

Muthukumar V, Melcher U, Pierce M, Wiley GB, Roe BA, Palmer MW, Thapa V, Ali A, Ding T (2009) Non-cultivated plants of the Tallgrass Prairie Preserve of northeastern Oklahoma frequently contain virus-like sequences in particulate fractions. *Virus Res.* **141**, 169-173. https://doi.org/10.1016/j.virusres.2008.06.016

Naidu RA, Bottenberg H, Subrahmanyam P, Kimmins FM, Robinson DJ, Thresh JM (1998) Epidemiology of groundnut rosette virus disease: current status and future research needs. *Ann. Appl. Biol.* **132**, 525-548. https://doi.org/10.1111/j.1744-7348.1998.tb05227.x

Natsuaki T, Muroi Y, Okuda S, Teranaka M (1990) Carrot temperate cryptoviruses and RNA-RNA hybridization among their double-stranded RNA segments. *Ann. Phytopath. Soc. Japan* **56**, 354-358. https://www.jstage.jst.go.jp/article/jjphytopath1918/56/3/56_3_354/_pdf

Nibert ML, Pyle JD, Firth AE (2016) A +1 ribosomal frameshifting motif prevalent among plant amalgaviruses. *Virology* **498**, 201-208. https://doi.org/10.1016/j.virol.2016.07.002

Petruzzello M (2019) List of plants in the family Apiaceae. *Encyclopedia Britannica*. https://www.britannica.com/topic/list-of-plants-in-the-family-Apiaceae-2038061. Accessed 30 October 2022

Quito-Avila DF, Jelkmann W, Tzanetakis IE, Keller K, Martin RR (2011) Complete sequence and genetic characterization of Raspberry latent virus, a novel member of the family Reoviridae. *Virus Res.* **155,** 397-405. https://doi: 10.1016/j.virusres.2010.11.008

Rivarez MPS, Pecman A, Bačnik K, Maksimović Carvalho Ferreira O, Vučurović A, Seljak G, Mehle N, Gutiérrez-Aguirre I, Ravnikar M, Kutnjak D (2022) In-depth study of tomato and weed viromes reveals undiscovered plant virus diversity in an agroecosystem. *bioRxiv*. https://doi.org/10.1101/2022.06.30.498278

Roossinck MJ (2012) Plant virus metagenomics: Biodiversity and ecology. *Ann. Rev. Genetics* **46**, 359-369. https://doi.org/10.1146/annurev-genet-110711-155600

Roossinck MJ, García-Arenal F (2015) Ecosystem simplification, biodiversity loss and plant virus emergence. *Curr. Opin. Virol.* **10**, 56–62. https://doi.org/10.1016/j.coviro.2015.01.005

Ryabov EV, Taliansky ME, Robinson DJ, Waterhouse PM, Murant AF, de Zoeten GA, Falk BW, Vetten HJ, Gibbs MJ (2012) Umbravirus. In: Virus taxonomy - Ninth report of the International Committee on Taxonomy of Viruses, King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (eds), Elsevier Academic Press, London (UK), pp 1191-1195.

Schönegger D, Babalola BM, Marais A, Faure C, Candresse T (2022a) Diversity of polerovirus-associated RNAs in the virome of wild and cultivated carrots. *Plant Pathology*, Online first. https://doi.org/10.1111/ppa.13623

Schönegger D, Marais A, Faure C, Candresse T (2022b) A new flavi-like virus identified in populations of wild carrots. *Arch Virol*. **167,** 2407-2409. https://doi: 10.1007/s00705-022-05544-1.

Stubbs LL (1948) A new virus disease of carrots: Its transmission, host range, and control. *Austr. J. Biol. Sci.* **1**,303-332. https://doi.org/10.1071/BI9480303

Susi H, Filloux D, Frilander MJ, Roumagnac P, Laine AL (2019) Diverse and variable virus communities in wild plant populations revealed by metagenomic tools. *PeerJ* **7**, e6140. https://doi.org/10.7717/peerj.6140

Svanella-DumasL, Tsarmpopoulos I, Marais A, Theil S, Faure C, Gaudin J, Candresse T (2018) Complete genome sequence of lettuce chordovirus 1 isolated from cultivated lettuce in France. *Arch. Virol*. **163**, 2543-2545. https://doi.org/10.1007/s00705-018-3858-y

Tan ST, Liu F, Lv J, Liu QL, Luo HM, Xu Y, Ma Y, Chen XJ, Lan PX, Chen HR, Cao MJ, Li F. (2021) Identification of two novel poleroviruses and the occurrence of Tobacco bushy top disease causal agents in natural plants. *Sci Rep*. **26**, 21045. https://doi: 10.1038/s41598-021-99320-x

Thapa V, McGlinn DJ, Melcher U, Palmer MW, Roossinck MJ (2015) Determinants of taxonomic composition of plant viruses at the Nature Conservancy's Tallgrass Prairie Preserve, Oklahoma. *Virus Evol.* **1**, 1-8. https://doi.org/10.1093/ve/vev007

Thompson JD, Higgins DG, Gibson TJ (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*. **11**, 4673-4680. https://doi: 10.1093/nar/22.22.4673

Tzanetakis I, Postman J, Martin R (2007) Identification, detection and transmission of a new vitivirus from Mentha. *Arch. Virol.* **152**, 2027-2033. https://doi.org/10.1007/s00705-007-1030-1

van Dijk P, Bos L (1985) Viral dieback of carrot and other Umbelliferae caused by the Anthriscus strain of parsnip yellow fleck virus, and its distinction from carrot motley dwarf. *Neth. J. Plant Pathol*. **91**, 169-187. https://doi.org/10.1007/BF02009678

Willenborg J, Menzel W, Vetten HJ, Maiss E (2009) Molecular characterization of two alphacryptovirus dsRNAs isolated from Daucus carota. *Arch. Virol.* **154,** 541-543. https://doi.org/10.1007/s00705-009-0314-z

Wu H, Fu P, Fu Q, Yu F, Peng Y (2022) Plant Virus Database: a resource for exploring the diversity of plant viruses and their interactions with hosts. *bioRxiv*. https://doi.org/10.1101/2022.03.20.485054

Yamashita S, Ohki S, Doi Y, Yora K (1976) Identification of two viruses associated with the carrot yellow leaf syndrome. *Ann. Phytopathol. Soc. Jpn*. **42**, 382-383.

Yoshida N (2020) Biological and genetic characterization of carrot red leaf virus and its associated virus/RNA isolated from carrots in Hokkaido, Japan, *Plant Pathol.* **69**, 1379-1389. https://doi.org/10.1111/ppa.13202

Table 1. Novel viruses identified in different carrot populations in France and Spain with details on the corresponding sequence library

Virus Family	Virus genus	virus name	virus acronym	Genome type	Accession number	scaffold length (nt)	mapped reads	average coverage	population type	country
Solemoviridae	*Polerovirus*	Carrot polerovirus 1	CaPV1	ssRNA+		5671	43652 (0.48%)	873x	wild*	France
Solemoviridae	*Polerovirus*	Carrot polerovirus 2	CaPV2			5213	14916 (0.15%)	324x	wild*	France
Solemoviridae	*Enamovirus*	Carrot enamovirus 1	CaEV1			5100	1403 (0.01%)	31x	wild	France
Tombusviridae	*Umbravirus*	Carrot umbravirus 1	CaUV1			3308	88394 (0.63%)	3033	wild	France
Tombusviridae	*Umbravirus*	Carrot umbravirus 2	CaUV2			4255	22748 (0.24%)	605x	wild	France
Tombusviridae	*Umbravirus*	Carrot umbravirus 3	CaUV3			4059	50896 (0.36%)	1421x	wild	France
Closteroviridae	*Closterovirus*	Carrot closterovirus 2	CtCV2			16302	3553362 (48.76%)	24536x	cultivated	France
Closteroviridae	*Closterovirus*	Carrot closterovirus 3	CtCV3			16291	602434 (11.26%)	4173x	cultivated	France
Betaflexiviridae	*Chordovirus*	Carrot chordovirus 3	CChV3			6881	2839 (0.05%)	39x	off-type	France
Betaflexiviridae	*Chordovirus*	Carrot chordovirus 4	CChV4			7703	1672 (0.02%)	23x	wild	France
Betaflexiviridae	*Vitivirus*	Carrot vitivirus 1	CaVV1			7546	14828 (0.43%)	221x	wild	France
Endornaviridae	*Alphaendornavirus*	Carrot alphaendornavirus	CaAEV1			14578	80742 (1.87)	608x	wild	France
linear ss satellite RNAs	unclassified	Carrot mottle virus satellite RNA 2	CtMoVsatRNA2			323	1777 9 (0.01%)	620x	wild	France
tombusvirus-like-associated RNA	unclassified	Carrot red leaf virus-associated RNA 2	CtRLVaRNA2		ON603907	Schönegger et al., 2022a				
small circular ss satellite RNAs	unclassified	Carrot red leaf virus satellite 1	CtRLV Sat1		OM962993	Babalola et al., 2022				
small circular ss satellite RNAs	unclassified	Carrot red leaf virus satellite 2	CtRLV Sat2		OM962994.1	Babalola et al., 2022				
Partitiviridae	*Deltapartitivirus*	Carrot cryptic virus 2	CaCV2	dsRNA		3329	932448 (24.19%)	31592x	wild	France
Partitiviridae	*Deltapartitivirus*	Carot cryptic virus 3	CaCV3			3345	408353 (14.99%)	13634x	wild	Spain
Amalgaviridae	*Amalgavirus*	Carrot amalgavirus 1	CaAV1			3223	26411 (0.26%)	929x	wild	France
Amalgaviridae	*Amalgavirus*	Carrot amalgavirus 2	CaAV2			777	130 (0.00%)	19x	wild	Spain
Reoviridae	*Reovirus*	Carrot reovirus 1	CaRV1			26081	2792566 (32.23%)	12089x	wild	France
Reoviridae	*Reovirus*	Carrot reovirus 2	CaRV2			3163	4372 (0.10%)	156x	wild	France
Totiviridae	*Totivirus*	Carrot-associated toti-like virus	CaaTLV			5128	403094 (4.98%)	8876x	wild	France
putative novel Riboviria	putative novel riboviria	Carrot flavi-like virus 1	CaFLV1	ssRNA?	OM681407.1	Schönegger et al., 2022b				
putative novel Riboviria	putative novel riboviria	Carrot Ker-like virus	CaKLV	ssRNA?		to be published elsewhere (TC, AM, DS et al., manuscript in preparation)				

Table 2: Known viruses identified in the carrot virome

Virus Family	Virus genus	Virus name	Acronym	Genome nucleic acid	Accession number
Solemoviridae	*Polerovirus*	Carrot red leaf virus	CtRLV	+ ssRNA	NC_006265
Solemoviridae	*Polerovirus*	Wild carrot red leaf virus	WCtRLV		LT615231
Tombusviridae	*Umbravirus*	Carrot mottle virus	CMoV		KF533714
Tombusviridae	*Umbravirus*	Wild carrot mottle virus	WCMoV		LT615232
Tombusviridae	*Umbravirus*	Carrot mottle mimic virus	CMoMV		NC_001726
Closteroviridae	*Closterovirus*	Carrot closterovirus 1	CtCV1		KF533697
Betaflexiviridae	*Closterovirus*	Carrot yellow leaf virus	CYLV		KF533699
Betaflexiviridae	*Chordovirus*	Carrot Chordovirus 1	CChoV1		NC_025469
Betaflexiviridae	*Chordovirus*	Carrot Chordovirus 2	CChoV2		NC_025468
Secoviridae	*Torradovirus*	Carrot Torradovirus 1	CaTV1		NC_025479 and NC_025480
Potyviridae	*Potyvirus*	Apium virus Y	AVY		NC_014905
Bromoviridae	*Ilarvirus*	Solanum nigrum ilarvirus 1	SnIV1		MN216370
linear ss satellite RNAs	*unclassified*	Carrot mottle virus satellite RNA	CMoVsatRNA		NC_030649
Tombusvirus like associated RNA	*unclassified*	Carrot red leaf virus associated RNA	CtRLVaRNA		NC_003871
Tombusvirus like associated RNA	*unclassified*	Arracacha latent virus E associated RNA	ALVEaRNA		MF136436
Tombusvirus like associated RNA	*unclassified*	Beet western yellows virus associated RNA	BWYVaRNA		KF533709
Umbra-like associated RNA	*unclassified*	Parsley umbravirus 1	ParUV1		OM419177.1
Partitiviridae	*Alphapartitivirus*	Carrot cryptic virus	CaCV1	dsRNA	NC_038824 and NC_038823
Partitiviridae	*Betapartitivirus*	Dill cryptic virus 2	DiCV2		NC_021147 and NC_021148
putative novel Riboviria	*putative novel riboviria*	Carrot-associated RNA virus 1	CaRNAV1	ssRNA?	

Figure 1. Geographical origin of the sampled carrot populations used in the present study (A-B) and proportion of known and novel viruses identified in their virome, together with the corresponding viral families (C). The maps show sampling locations in France (A) over two years (2019 and 2020) and in Spain (B) in 2021, with green pins corresponding to wild carrot populations and red and orange pins to cultivated and off-type carrot populations respectively. Maps were created using framacarte.org

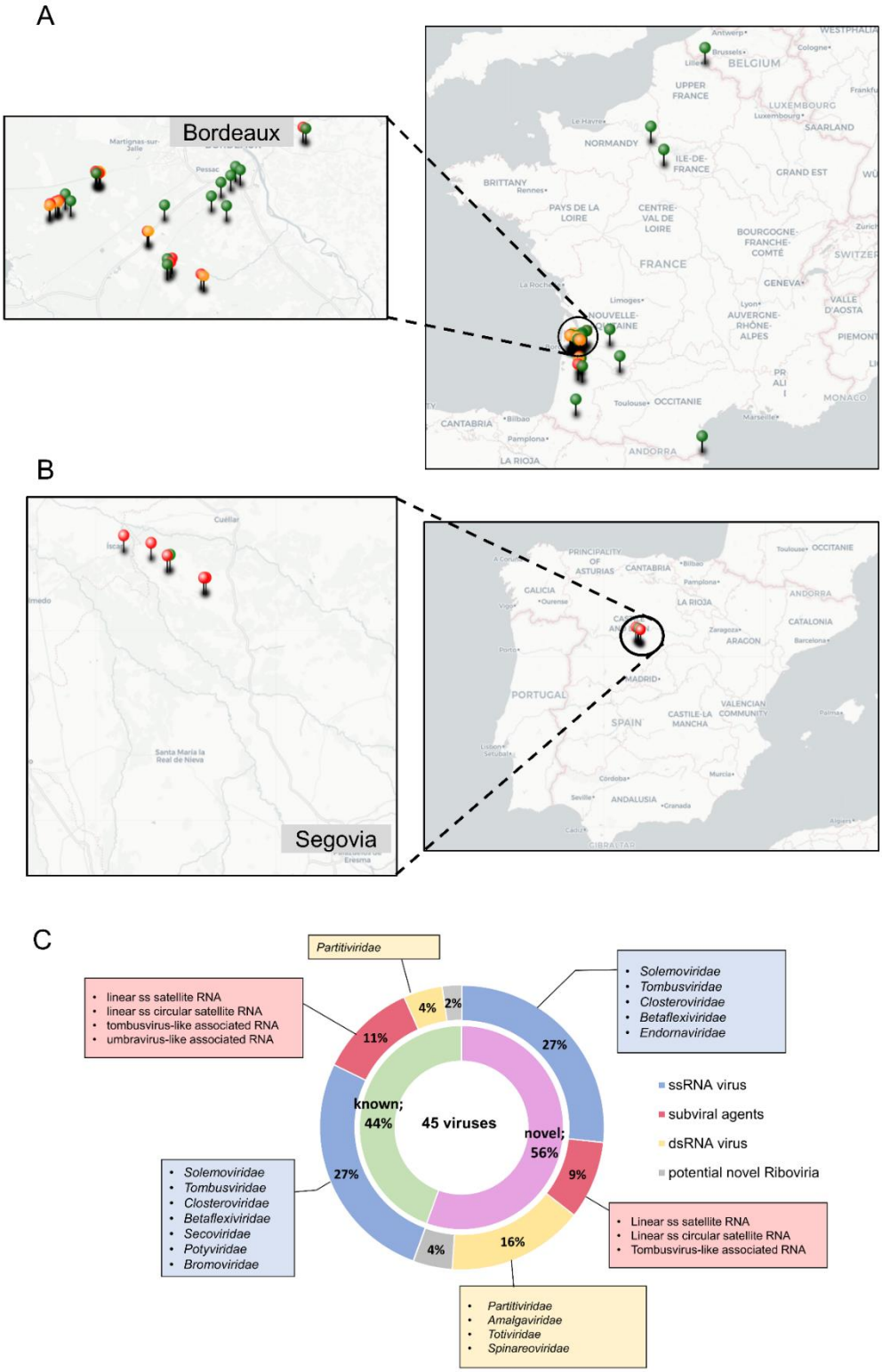


Figure 2. Neighbour-joining trees reconstructed from full-length nucleotide alignments of the RdRp gene from identified umbraviruses and related reference sequences from GenBank (A) and from partial RdRp gene sequences from umbra-like-associated RNAs (B). A strict nucleotide identity distance was used and bootstrap values above 70% are shown (1,000 replicates). The scale bars represent 10% (A) and 5% (B) nucleotide divergence between isolates. Accession numbers are indicated for each reference sequence and novel viruses are indicated by black diamonds. The corresponding genus or umbra-like-associated RNA class is indicated to the right of each bracket.

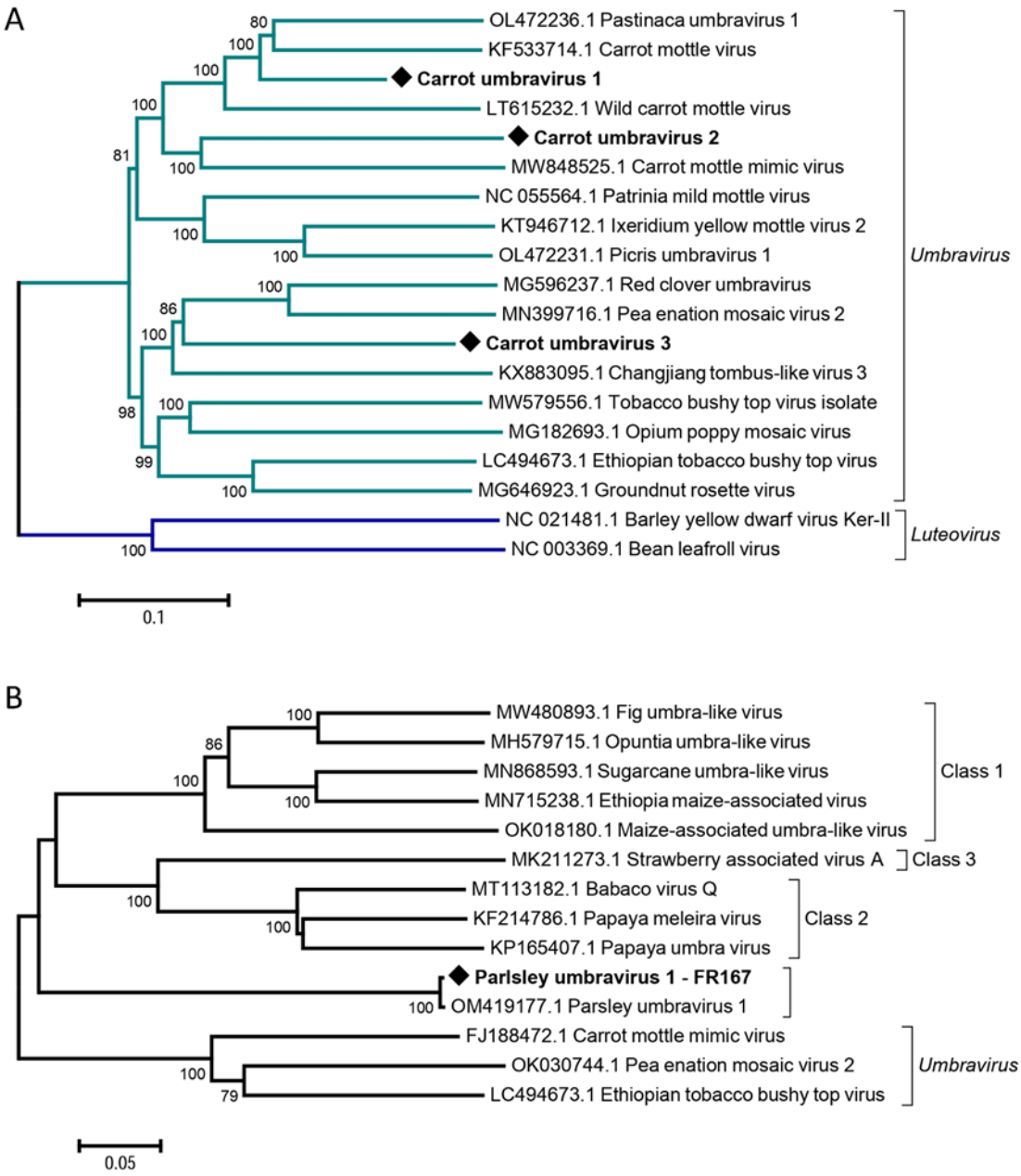


Figure 3: Schematic representation of the genome organization of the novel viruses carrot enamovirus 1 (CaEV1) and carrot polerovirus 1 (CaPV1). The genomes of pea enation mosaic virus 1 (PEMV1, *Enamovirus*, GenBank MN497824) and of carrot red leaf virus (CtRLV, *Polerovirus*, GenBank LC434063) are presented for comparison. Open reading frames are indicated by differently coloured arrows and their protein function described in the colour code.

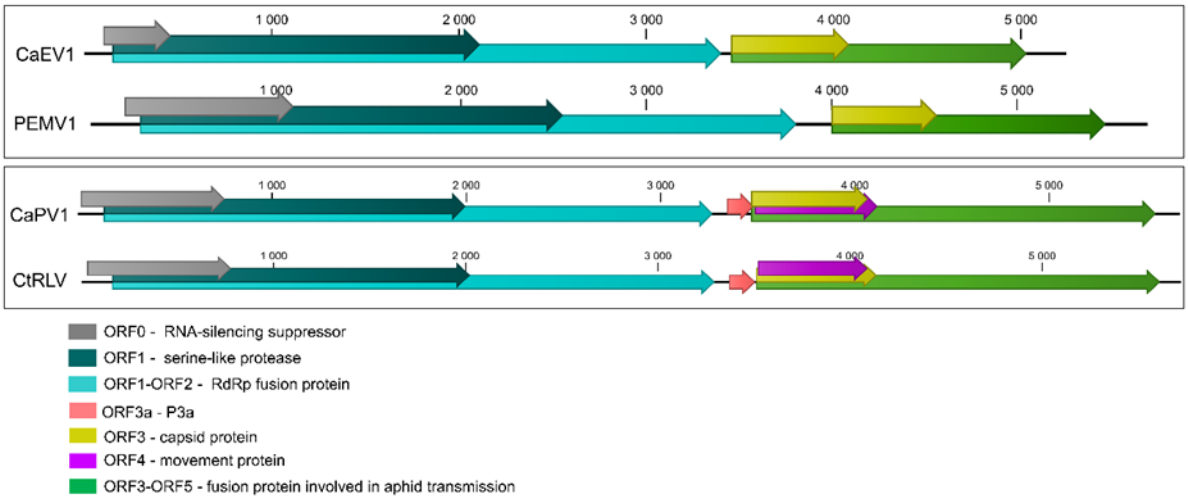


Figure 4. Neighbour-joining tree reconstructed from amino acid alignment of the full length RNA-dependent RNA polymerase from known and novel members of the *Polerovirus* and *Enamovirus* genera identified and of related reference sequences from GenBank. A strict amino acid identity distance was used and bootstrap values above 70% are shown (1,000 replicates). The scale bar represents 10% amino acid divergence between sequences. Accession numbers are indicated for each reference sequence and novel viruses are indicated by black diamonds. The corresponding genus is indicated to the right of each bracket.

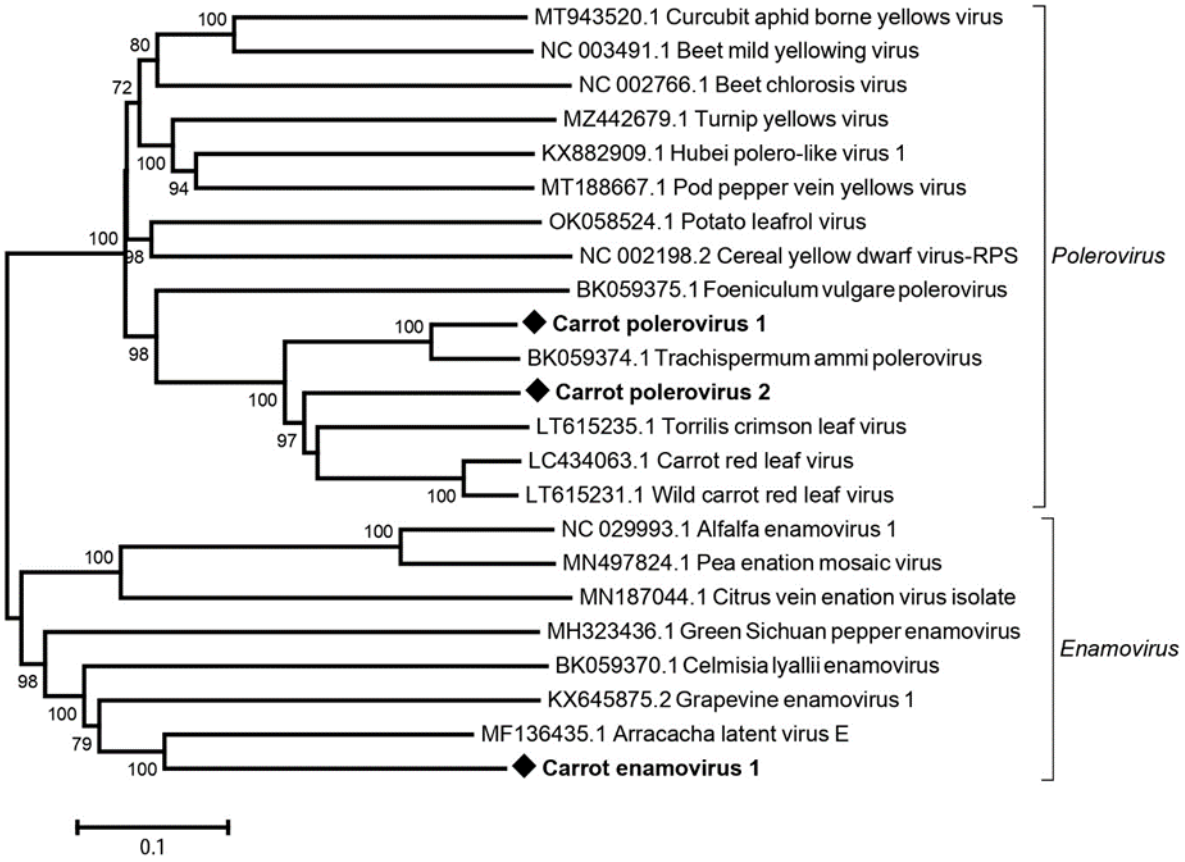


Figure 5. Neighbour-joining trees reconstructed from amino acid alignments of the RNA-dependent RNA polymerase (A) and the capsid protein of novel viruses in the family *Betaflexiviridae* identified in the French carrot virome, the most related reference sequences in GenBank and representative members of each genus. A strict amino acid identity distance was used and bootstrap values above 70% (1000x replicates) are shown. The scale bars represent 10% amino acid divergence between sequences. Novel viruses are indicated by black diamonds and accession numbers are given for each reference sequence. Genera are indicated in italics at the right of the trees.

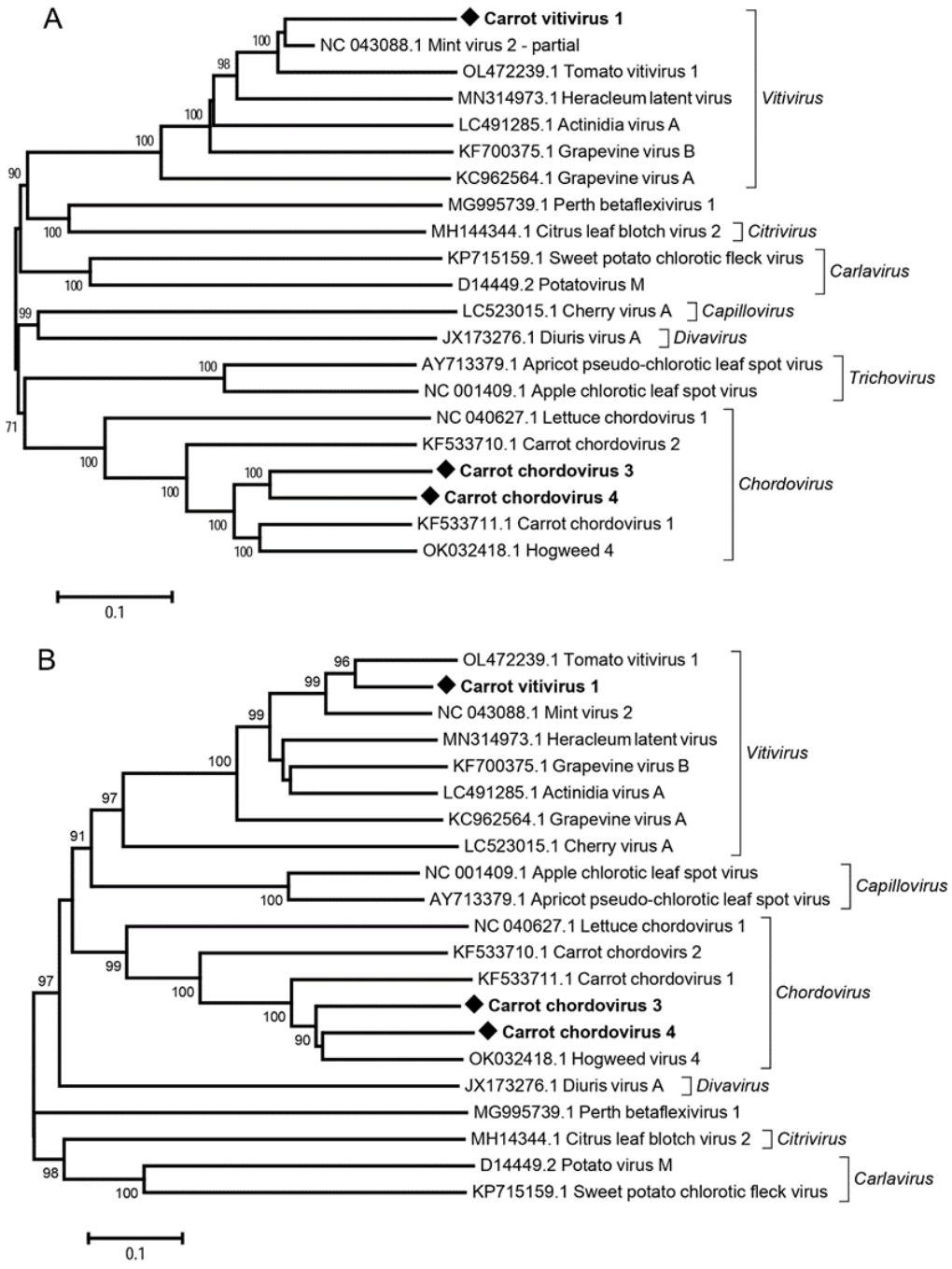


Figure 6. Schematic representation of the genome organization of selected chordoviruses (A) and vitiviruses (B). Viruses included in (A) are the novel carrot chordoviruses 3 and 4 (CChoV 3 and 4) and lettuce chordovirus 1 (LeCV1, GenBank NC_040627). Viruses included in (B) are the novel carrot vitivirus 1 (CaVV1), grapevine virus A (GVA, GenBank KC962564) and tomato vitivirus 1 (ToVV1, GenBank OL472239). Missing sequences in the novel genomes reported here are indicated by Ns at the corresponding genome location. Predicted open reading frames are indicated by arrows of different colours and the function of the encoded protein given in the colour code.

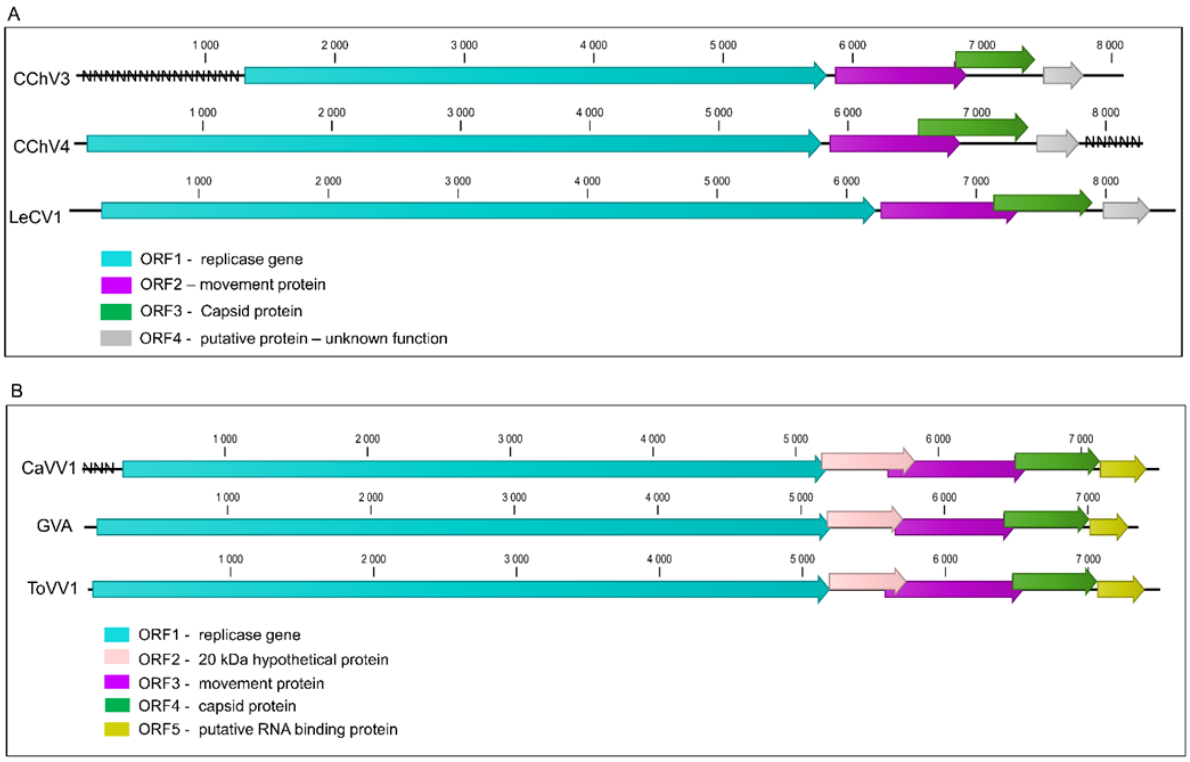


Figure 7. Schematic representation of the genomic organization of known and novel *Closterovirus* genus members. Viruses included are citrus tristeza virus (CTV), carrot closterovirus 1 (CtCV1), carrot yellow leaf virus (CYLV), and the novel carrot closteroviruses 2 and 3 (CtCV2 and CtCV3). ORFs are indicated as arrows in different colours and their function described in the colour code. Vertical arrows indicate putative +1 ribosomal frameshifts (+1RFS).

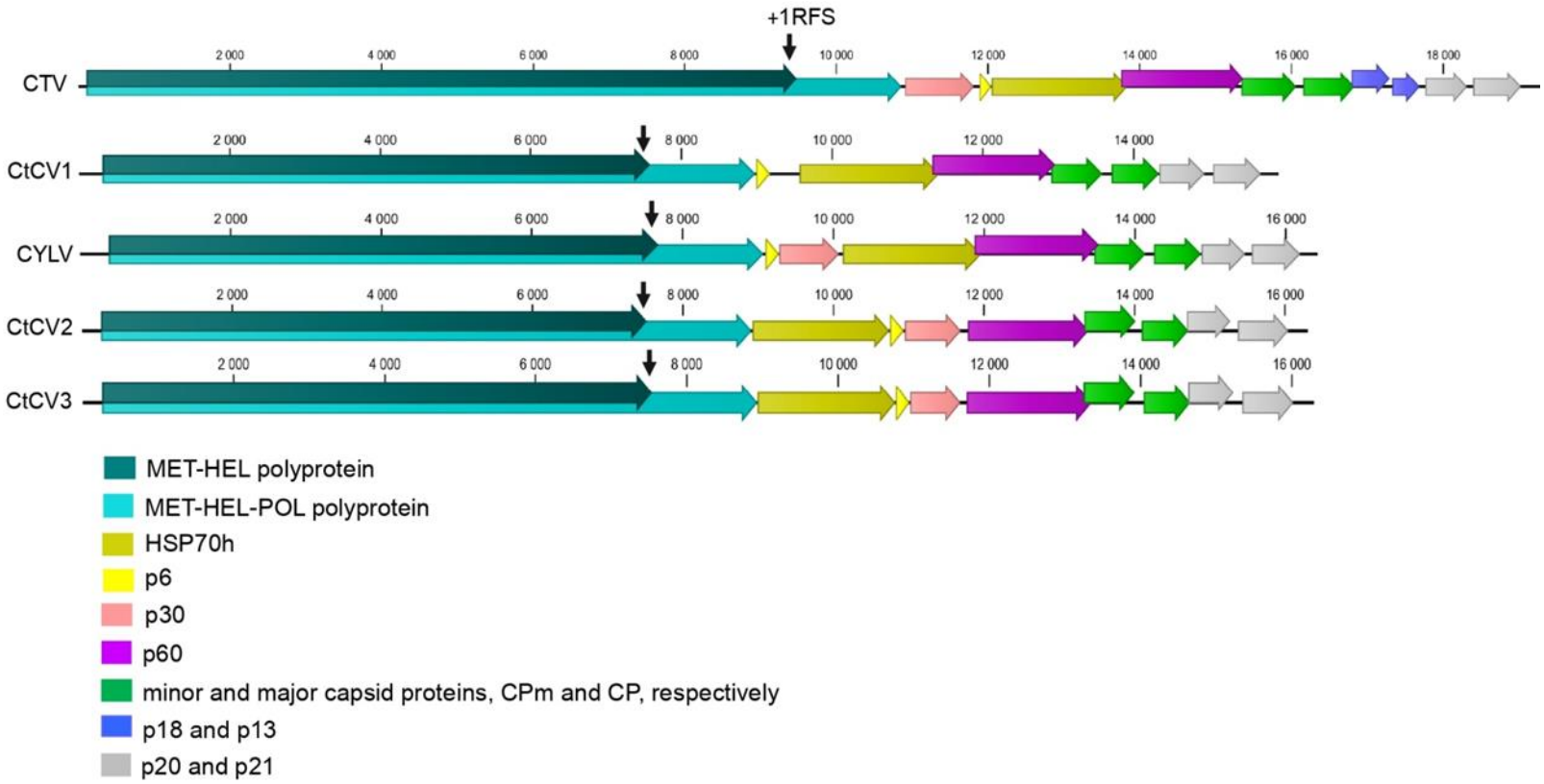


Figure 8. **Neighbour-joining tree reconstructed from the amino acid alignment of the RNA-dependent RNA polymerase of the novel carrot alphaendornavirus 1 and of reference sequences of members in the *Alphaendornavirus* and *Betaendornavirus* genera retrieved from GenBank**. A strict amino acid identity distance was used and bootstrap values above 70% (1,000x replicates) are shown. The scale bar represents 10% amino acid divergence between sequences. Carrot alphaendornavirus 1 is indicated by a black diamond. Accession numbers are given for each reference sequence. Genera are indicated in italics on the right of lateral brackets

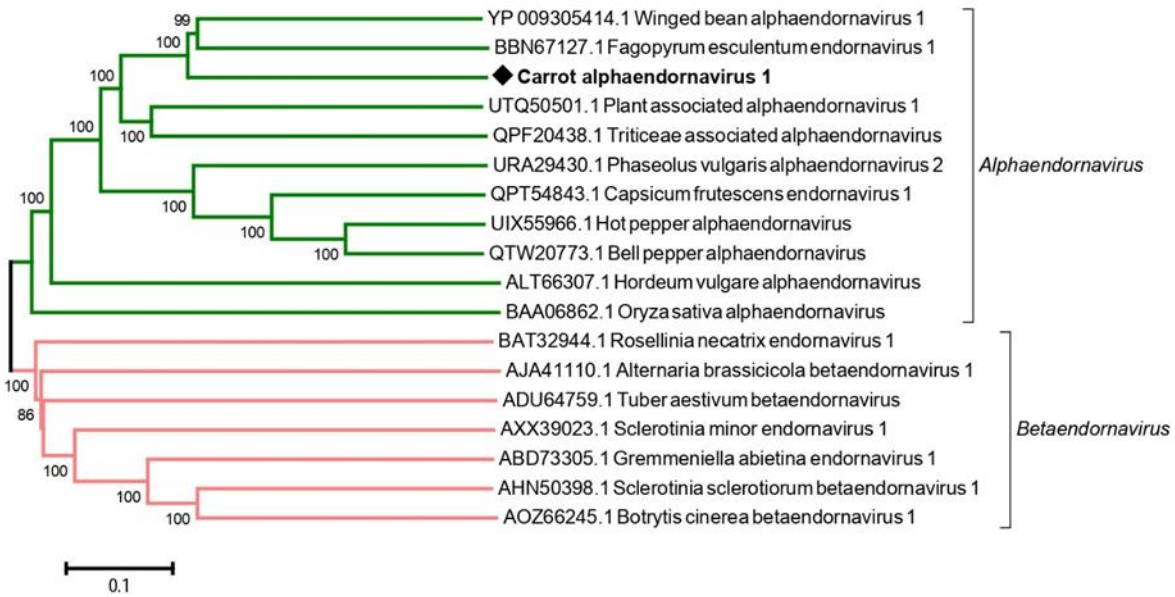


Figure 9. Neighbour-joining tree reconstructed from amino acid alignment of the RNA-dependent RNA polymerase of the newly identified carrot reovirus 1 and carrot reovirus 2 and of representative *Spinareoviridae* members and related unassigned viruses retrieved from GenBank. A strict amino acid distance was used and bootstrap values above 70% (1000 replicates) are shown. The scale bar represents 10% amino acid divergence between sequences. Novel viruses identified in this study are indicated by black diamonds and accession numbers are shown for each reference sequence. Genera are shown on the right of lateral brackets.

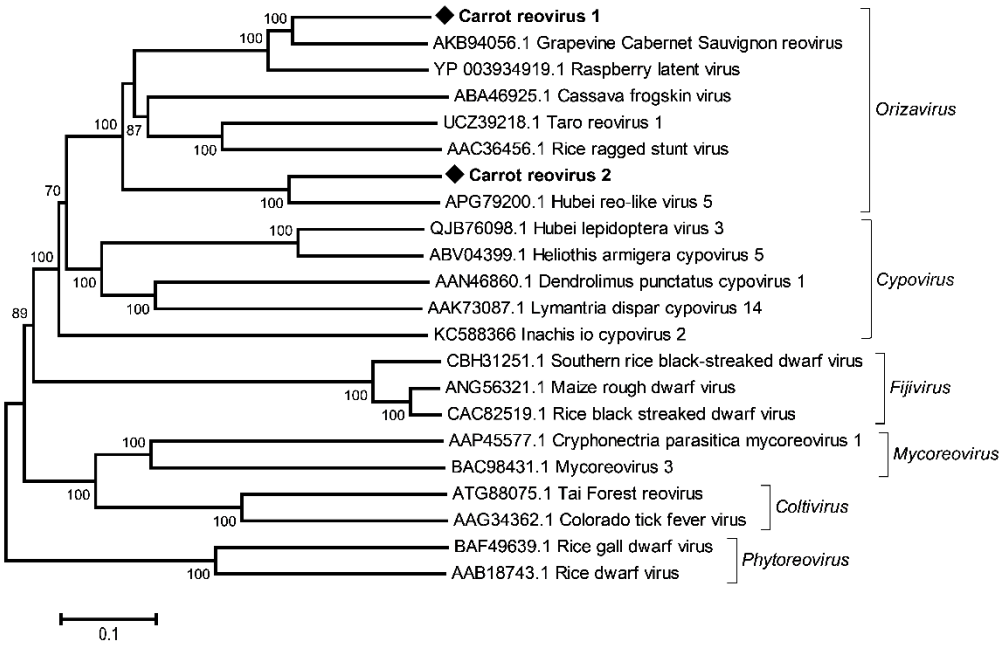


Figure 10. Neighbour-joining trees reconstructed from amino acid alignments of RNA-dependent RNA polymerase sequences of *Partitiviridae* members (A), *Amalgaviridae* members (B), and *Totiviridae* members (C). A strict amino acid identity distance was used and bootstrap values above 70% (1,000x replicates) are shown. Scale bars represent 10% amino acid divergence between sequences. Colour coding denotes the different genera within a family or, for totiviruses, the host (green for plant-associated viruses, brown for fungi-associated viruses). Novel viruses identified in this study are indicated by black diamonds and accession numbers are shown for each reference sequence. Genera are shown on the right of lateral brackets.

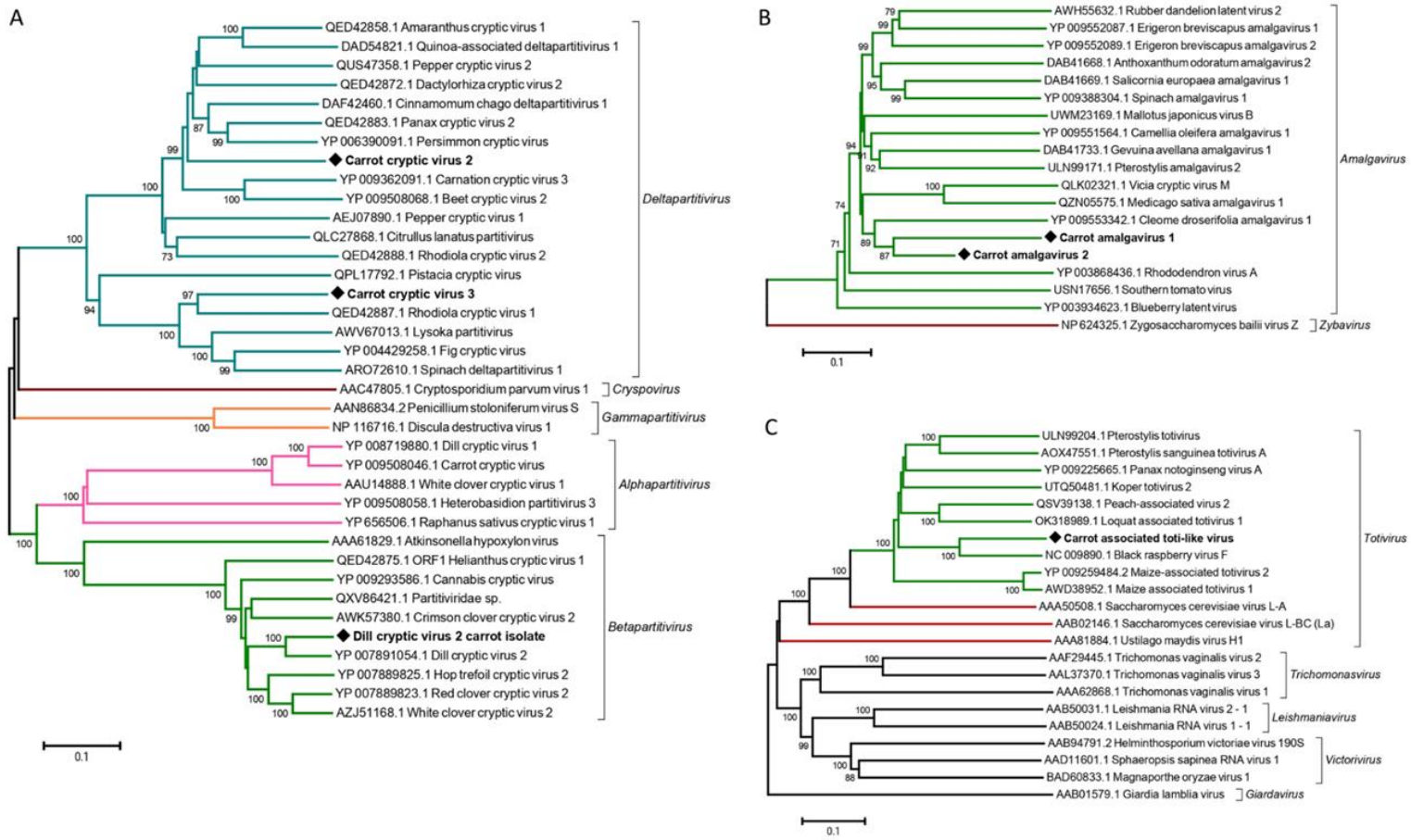
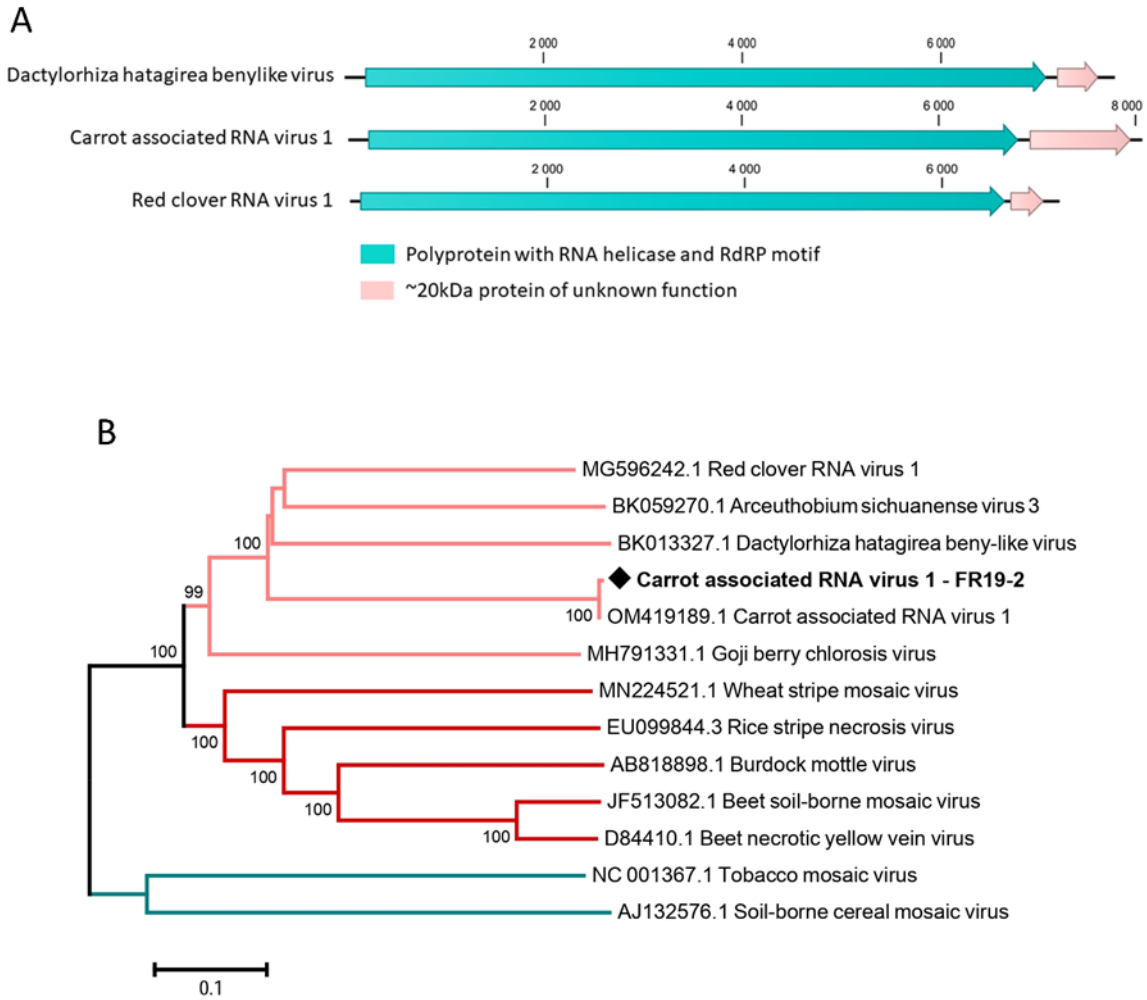


Figure 11: **Genomic organization of unclassified novel viruses with homology to the *Benyviridae* family, including carrot-associated RNA virus 1 (A) and neighbour-joining tree reconstructed from amino acid alignment of the RNA-dependent RNA polymerase of carrot-associated RNA virus 1 and of closely related sequences from GenBank (clade indicated in light pink), of selected *Benyviridae* members (clade indicated in red) and of selected *Virgaviridae* members (clade indicated in blue) (B)**. Carrot-associated RNA virus 1 is indicated by a black diamond and accession numbers are given for each reference sequence. A strict amino acid identity distance was used and bootstrap values above 70% (1,000 replicates) are shown. The scale bar represents 10% amino acid divergence between sequences.



Supplementay Materials [all Supplementary Tables are available at https://doi.org/10.57745/HRBVNR, the one that fit into A4 format is also provided here]

Supplementary Tables

Table S1: information on sampled carrot populations in France and Spain

Table S2: RT-PCR and RACE primer pairs targenting known and novel viruses in the carrot virome

Table S3: Umbravirus RdRp gene and complete genome pairwise nucleotide

Table S4: Solemoviridae RdRp pairwise amino acid identities

Table S5: Geographical distribution of the various viruses identified in the carrot virome

Table S6: Betaflexiviridae RdRp pairwise amino acid identities

Table S7: Closteroviridae RdRp, CP and HSP70h pairwise amino acid identities

Table S8: Endornaviridae complete genome pairwise nucleotide identities and RdRp

Table S9: Spinareoviridae RdRp pairwise amino acid identities

Table S10: Partitiviriade RdRp and CP pairwise amino acid identities

Table S11: Amalgaviridae RdRp fusion protein pairwise amino acid identities

Table S12: Totiviridae RdRp and CP genes pairwise amino acid identities

Table S13: Benyviridae RdRp gene pairwise amino acid identities

Supplementary Figures

Supplementary Figure S1: Recombination events identified for members of the Umbravirus and Polerovirus genus

Supplementary Table S5

		Country	France						Spain
		sampling year	2019	2020					2021
		Region	Nouvelle Aquitaine	Nouvelle Aquitaine	Île-de-France	Hauts-de-France	Normandy	Occitanie	Castile-León/Segovia
Virus family	Virus genus	number of sampled carrot populations	16	21	2	1	3	1	6
Solemoviridae	*Polerovirus*	Carrot red leaf virus							
		Wild carrot red leaf virus							
		Carrot polerovirus 1							
		Carrot polerovirus 2							
	Enamovirus	Carrot enamovirus 1							
Tombusviridae	*Umbravirus*	Carrot mottle virus							
		Wild carrot mottle virus							
		Carrot mottle mimic virus							
		Carrot umbravirus 1							
		Carrot umbravirus 2							
		Carrot umbravirus 3							
unclassified subviral agents	Umbravirus-like-associated RNA	Parsley umbravirus 1							
	linear ss satellite RNA	Carrot mottle virus satellite RNA							
		Carrot mottle virus satellite RNA 2							
	Tombusvirus-like associate RNAs	Carrot red leaf virus-associated RNA							
		Arracacha latent virus E-associated RNA							
		Carrot red leaf virus-associated RNA 2							
		Beet western yellows virus-associated RNA							
	small circular ss satellite RNAs	Carrot red leaf virus Satellite 1							
		Carrot red leaf virus Satellite 2							
Closteroviridae	*Closterovirus*	Carrot yellow leaf virus							
		Carrot closterovirus 1							
		Carrot closterovirus 2							
		Carrot closterovirus 3							
Betaflexiviridae	*Chordovirus*	Carrot chordovirus 1							
		Carrot chordovirus 2							
		Carrot chordovirus 3							
		Carrot chordovirus 4							
	Vitivirus	Carrot vitivirus 1							
Potyviridae	*Potyvirus*	Apium virus Y							
Secoviridae	*Torradovirus*	Carrot torradovirus 1							
Bromoviridae	*Ilarvirus*	Solanum nigrum ilarvirus 1							
Partitiviridae	*Alphapartitivirus*	Carrot cryptic virus							
	Betapartitivirus	Dill cryptiv virus 2 - carrot isolate							
	Deltapartitivirus	Carrot cryptic virus 2							
		Carrot cryptic virus 3							
Reoviridae	*Reovirus*	Carrot reovirus 1							
		Carrot reovirus 2							
Endornaviridae	*Alphaendornavirus*	Carrot endornavirus 1							
Amalgaviridae	*Amalgavirus*	Carrot amalgavirus 1							
		Carrot amalgavirus 2							
unclassified Riboviria	unclassified	Carrot-associated toti-like virus							
		Carrot-associated RNA virus 1							
		Carrot flavi-like virus							
		Carrot Ker-like virus							

Figure S1

Umbraviruses

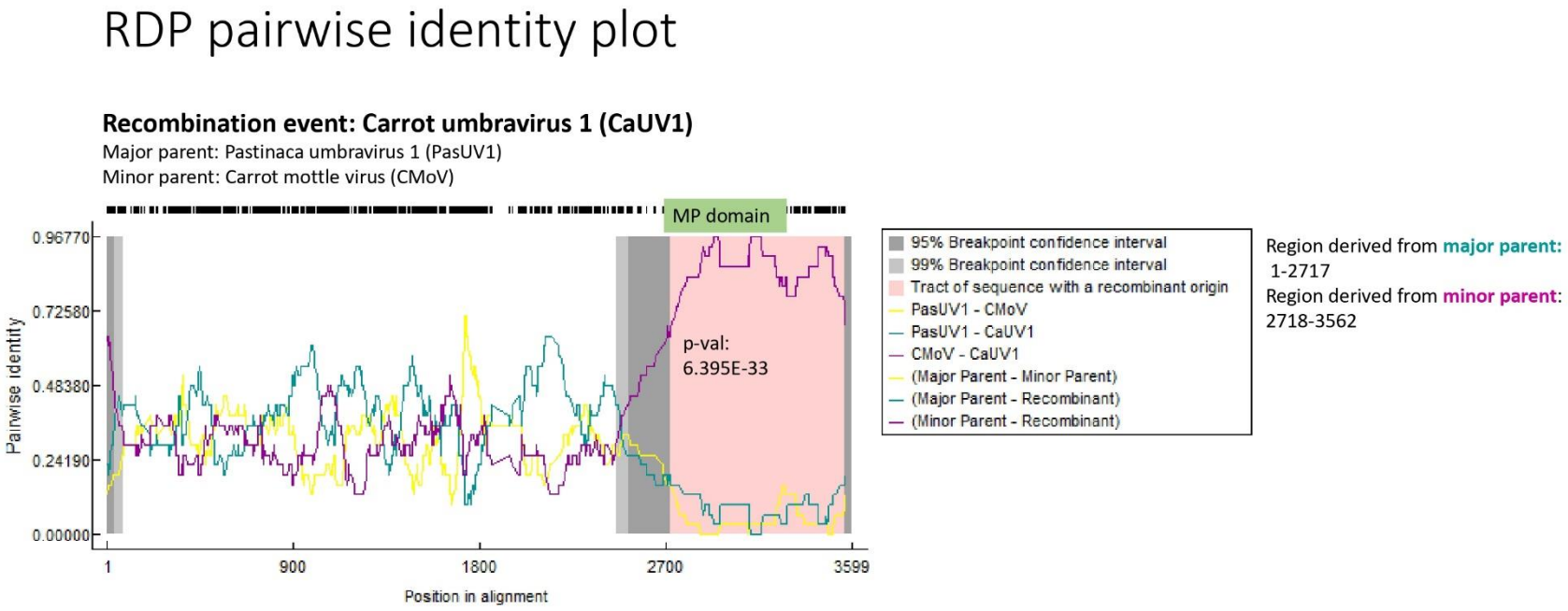


Figure S1 (continued)

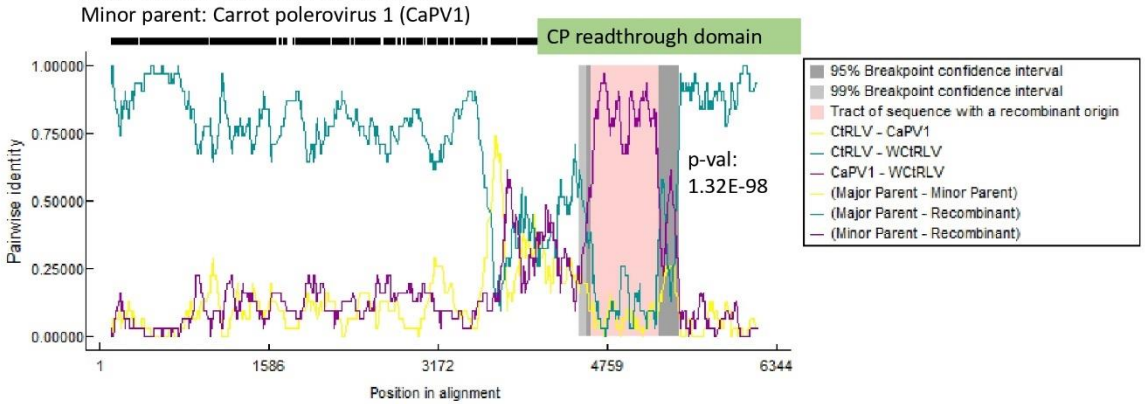
Polerovirus

RDP pairwise identity plot

Recombination event: Wild carrot red leaf virus (WCtRLV)

Major parent: Carrot red leaf virus (CtRLV)

Minor parent: Carrot polerovirus 1 (CaPV1)



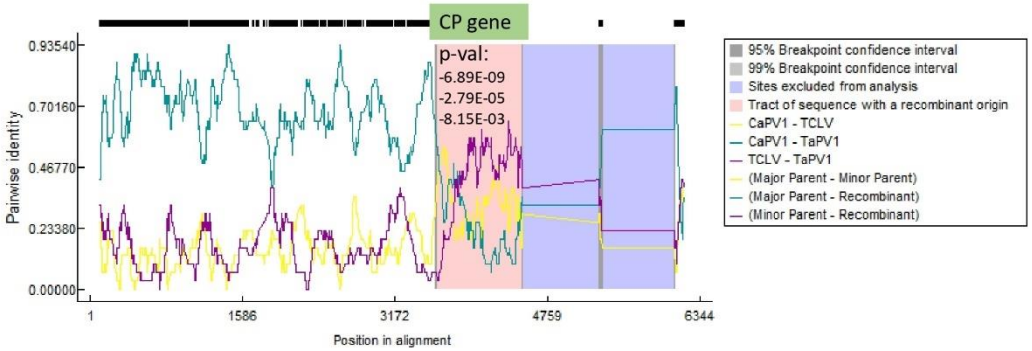
Region derived from **major parent:** 1-4601 and 5239-6344

Region derived from **minor parent**: 4602-5238

Recombination event: Trachispermum ammi polerovirus 1

Major parent: Carrot Polerovirus 1

Minor parent: Torrilis crimson leaf virus



Region derived from **major parent:** 1-3617 and 4498-6344

Region derived from **minor parent**: 3618-4497

A new flavi-like virus identified in populations of wild carrots

Diversity of polerovirus-associated RNAs in the virome of wild and cultivated carrots

CHAPTER III

Metagenomic analysis of cultivated and wild carrots reveals different virome fingerprints and provides insight into viral fluxes between carrot populations

Metagenomic analysis of cultivated and wild carrots reveals different virome fingerprints and provides insight into viral fluxes between carrot populations

Deborah Schönegger[1], Armelle Marais[1], Fernando Garcia Arenal[2], Marcos Martinez Jimenez[2], Michael McLeish[2], Marie Lefebvre[1], Chantal Faure[1], Laurence Svanella-Dumas[1], Thierry Candresse[1*]

[1]Univ. Bordeaux, INRAE, UMR BFP, CS20032, 33882, Villenave d'Ornon Cedex, France.

[2] Centro de Biotecnología y Genómica de Plantas (CBGP), Universidad Politécnica de Madrid (UPM) and E.T.S.I. Agronómica, Alimentaria y de Biosistemas, Campus de Montegancedo, UPM, 28223 Pozuelo de Alarcón, Madrid, Spain.

*** Author for correspondence:** thierry.candresse@inrae.fr Tel : (+33) 557 12 23 89

Abstract word count: 246 words

Text word count: 8482 words (Introduction-Materials and Methods-Results-Discussion)

Tables and Figures: 3 Tables, 6 Figures, 6 supplementary Tables and Figures

Running title: Contrasted virome fingerprints in wild and cultivated carrots

ABSTRACT

A double-stranded RNA-based high-throughput sequencing (HTS) approach was used to analyze the virome of 31 wild, cultivated and off-type carrot populations sampled in 2019-2020 in France. This revealed a rich virome comprising 36 viruses or virus-like agents of which 18 are novel. Hierarchical clustering and bipartite network analyses demonstrate surprisingly different virome fingerprints between wild and field (cultivated and off-type) carrots. This is in part due to differences in persistent viruses infection patterns but differences were also observed for many aphid-borne viruses. About half of the viruses are shared among all three carrot population types, while some agents were uniquely associated to either wild or field carrots. RT-PCR testing of individual plants from eight populations for 16 aphid-borne viruses indicated widely different prevalence levels between field and wild carrots for many viruses. Despite their different genetic makeup, off-type carrots had a virome that was closely related to that of cultivated carrots growing side by side with them, suggesting a major impact of agricultural growth conditions in shaping virome richness, composition and individual viruses prevalence. While amplicons and HTS contigs sequence analysis provided evidence for the existence of viral fluxes between the wild and field compartments for most aphid-transmitted viruses, population genetics analyses indicated the existence of barriers limiting or altogether preventing these fluxes for several viruses or for some lineages within other species. The bases for these barriers, which might reflect differential susceptibility of carrot populations, vector preferences or yet other factors remain to be identified.

Keywords: virome, dsRNA, high-throughput sequencing, reservoir, viral fluxes

INTRODUCTION

Viruses are widespread biological entities that depend on their host to maintain their life cycle and are often transmitted by vectors. Their role is mainly regarded as that of disease-causing agents in humans, animals and plants. However, in wild plants, viruses are often asymptomatic or associated with mild or indistinct abnormalities (Cooper and Jones, 2006), which may be due to a long co-evolution between plants and viruses (Lefeuvre *et al.*, 2019). In addition, a number of mechanisms are theoretically able to limit pathogens impact in wild plant communities, for example high plant species diversity with ensuing host dilution effects by non-host plant species, lower host population densities, and the presence of natural predators and parasites that may limit vector populations (Cooper and Jones, 2006).

Major milestones in history, such as the domestication of wild plant species and the advent of agriculture around 10,000 years ago, or more recently global warming and the globalisation of markets, have changed the dynamics between hosts, pathogens and their environment (Cooper and Jones, 2006; Jones, 2014). These changes have likely led to more encounters between otherwise separate hosts and pathogens, facilitating virus emergence and disease epidemics (Elena *et al.*, 2014). In modern agriculture, crops are usually part of a complex agro-ecosystem and lie adjacent to natural plant communities. This so called agro-ecological interface plays an important role in virus interactions between cultivated and wild environments (Haňcinský *et al.*, 2020). Wild plants can serve as reservoirs for viruses that can then spread to neighbouring crops and, conversely, wild plant communities are increasingly exposed to plant viruses that can spill over from agroecosystems (Roossinck and García-Arenal, 2015; Malmstrom and Alexander, 2016). Therefore, the wild and cultivated compartments may interact more or less strongly in ways that are expected to shape virus populations. The undisturbed interactions between plants, vectors, and viruses in natural environments allow viruses to vary and adapt to different hosts and vectors, making wild plants the main source of virus diversity (Cooper and

Jones, 2006). In addition, host biodiversity has been predicted to alter disease risk by either increasing (amplification effect) or decreasing (dilution effect) pathogen diversity and prevalence (Pagán *et al.*, 2012). To understand the ecology and composition of virus communities and their interaction and circulation between wild reservoirs and crop hosts, it is important to analyse virus communities across agro-ecological interfaces. Innovative genomic tools introduced in the 21st century such as high-throughput sequencing (HTS), allow to analyse in an unbiased way the structure of microbial communities in different ecosystems, including viral populations, allowing the detection of new viruses without any prior knowledge. During the last decade, a number of studies have been published looking at viral diversity in natural ecosystems (Muthukumar *et al.*, 2009; Roossinck *et al.*, 2010; Roosinck, 2011; Roossinck *et al.*, 2015; Thapa *et al.*, 2015; Fraile *et al.*, 2017; Bernardo *et al.*, 2018; Susi *et al.*, 2019, Mahlanza *et al.*, 2022) as well as virus movement between wild and cultivated hosts and the role of wild plants as virus reservoirs (Tugume *et al.*, 2016; Ma *et al.*, 2020, Maachi *et al.*, 2022; Rivarez *et al.*, 2022). Bernardo *et al.* (2018), using a landscape-scale geometagenomic approach, found that virus prevalence is higher in cultivated areas, with some virus families preferentially associated with agriculture. Similarly, Thapa *et al.* (2015) found a significant relationship between plant virus community composition and host identity. Ma *et al.* (2020) compared viromes of tomato and of populations of the wild *Solanum nigrum* and found that these related species share a relatively low proportion of Operational Taxonomic Units (OTUs, a proxy for viral species) and exhibit high variability in virome richness between sites. They also showed that there was crosstalk between the viral populations in populations of these two plants but that there were also some unexplained cases where crosstalk was limited. Moreover, Maachi *et al.* (2022) observed virome overlap exclusively between crops and weeds of the same family, but also showed differentiation of some viral populations, suggesting that at least for one virus, there were few exchanges between crop and weeds of the same family. It can

therefore be expected that intrinsic plant and virus characteristics strongly influence the movement of viruses at the agroecological interface. As highlighted by Moury *et al.* (2017), the taxonomic relatedness of plants is an important barrier to virus infection of new hosts and thus is a strong contributing factor to virus host range determination. Similarly, Sallinen *et al.* (2020) showed that the host genotype plays a key role in the composition of virus communities. The host genetics and biology are therefore expected to ultimately influence virus communities and their circulation in the environment. In addition, environmental factors such as plant growth conditions and soil nutrient status can impact pathogen community structure and affect disease dynamics (Susi and Laine, 2021). Particularly for viruses circulating through the action of vectors, changes in host density, especially at the agro-ecological interface, can alter feeding behaviour, leading to increased frequency of inoculum transmission and consequent disease incidence (Burdon and Chilvers, 1982).

Metagenomic studies aiming to compare virus communities, diversity and virus transfers between crops and weeds have so far focused on different species belonging to the same family or belonging to even more distantly related taxonomic groups (Ma *et al.*, 2020; Maachi *et al.*, 2022; Rivarez *et al.*, 2022). Cultivated (*Daucus carota* ssp. *sativus*) and wild (*Daucus carota* ssp. *carota*) carrots belong to the same species and represent a situation with the least botanical and genetic distance between a crop and a frequent wild weed that could act as a reservoir. Thus, wild and cultivated carrots represent a particularly interesting model system for comparative metagenomic analysis and for the analysis of viral fluxes between the wild and cultivated compartments. In particular, this pathosystem is expected to limit the impact of the species barrier and thus to impose fewer constraints on virus flow. Carrot is the most widely cultivated crop in the *Apiaceae* family and is grown on 1.1 million hectares worldwide (FAO, 2020). The history of carrot domestication is relatively short, and the first domestication events are thought to have occurred in southwestern Asia (Afghanistan) only about 1,100 years ago

(Rong *et al.*, 2014). The wild carrot, a biennial, weedy precursor to cultivated carrots is widely distributed in temperate regions and native to Europe, southwestern Asia and North Africa (Praciak, 2022). In France, the region Nouvelle-Aquitaine represents the largest carrot-growing area and 99% of the regional production area is concentrated in the Gironde and Landes departments (Anon, 2018), where carrot cultivation benefit from deep sandy soils. At the same time, these particular local pedoclimatic conditions have the consequence that wild carrots are in many places the only *Apiaceae* species growing outside of wet areas along streams and ditches.

In total, more than 30 viruses from 12 different families are known to infect carrots (Wu *et al.*, 2022; Mihara *et al.*, 2016; Brunt *et al.*, 1996). In particular, the families *Tombusviridae, Solemoviridae, Closteroviridae, Secoviridae* and *Potyviridae* contain well-known carrot-infecting viruses that can cause severe damage to carrot production (Adams *et al.*, 2014; Rozado-Aguirre *et al.*, 2016, Yoshida, 2020; Latham and Jones, 2004). For instance, Carrot motley dwarf disease (CMD) is a well-known carrot disease described for the first time by Stubbs (1948). It is caused by a mixed infection of a polerovirus, carrot red leaf virus (CtRLV) and an umbravirus, either carrot mottle virus (CMoV) or carrot mottle mimic virus (CMoMV) (Yoshida, 2020). In some instances, an associated satellite can also be involved in the complex (Menzel *et al.* 2009; Yoshida, 2020). The disease is known to be transmitted by the willow carrot aphid (*Cavariella aegopodii*) in a persistent and non-propagative way (Stubbs, 1948; Yoshida, 2020). However, it has been recently shown that the aphid *Myzus persicae* is also able to transmit CMD-involved viruses to different species of the *Apiaceae* family under laboratory conditions (Naseem *et al.*, 2016). Infected plants exhibit chlorotic mottling and yellowing of foliage and can suffer from severe stunting (Stubbs, 1948; Menzel, 2009, Yoshida, 2020).

As part of a large-scale virome study, different cultivated and wild carrot populations sampled along a north-south transect of France were analysed using an HTS-based approach. This

sequencing effort allowed the identification of a large number of known and new viruses. Genomic contigs have been assembled for all these agents (Schönegger et al., in preparation) and have been used here to analyse the virome structure of selected wild and cultivated carrot populations sampled over a two years period in Nouvelle-Aquitaine, France. The genetic structure of the populations of selected viruses was further analysed by RT-PCR amplicon sequencing and HTS virome analysis of individual plants. This study enables a holistic virome analysis at the agro-ecological interface with an in-depth insight into virus circulation between wild plants and a crop using carrots as a pathosystem characterized by particularly low taxonomic barriers to virus flow.

Expecting higher viral species richness and lower virus prevalence in wild compared to cultivated carrots, we specifically aimed to address several questions: (i) how different or similar are the viromes of cultivated and wild carrots? (ii) is there evidence of virus exchange between wild and cultivated carrot populations or, on the contrary, evidence for the existence of barriers to virus movement? (iii) are specific virus assemblages identified in relation to their life history traits in a complex viral system of multiple dependence and assistance interactions?

MATERIALS AND METHODS

Sample collection and processing

In summer 2019, 16 different carrot populations, including eight wild carrot populations, four cultivated populations and four off-type populations growing within cultivated carrot fields and showing some characteristics of wild carrots (early bolting, small, often white, tap root), suggesting an origin from contaminations of the planted commercial seed lots (thereafter referred to as off-type carrots) were sampled in the Landes and Gironde departments of Nouvelle-Aquitaine, France. A similar sampling was carried out in the summer of 2020 in the department of Gironde and included seven cultivated populations (two of which were respectively sampled in the middle and at the edge of the same field), three off-type populations

and five wild carrot populations (two of which grew at the edge of two of the sampled fields). Thus, a total of 31 carrot populations were sampled in Nouvelle-Aquitaine (France) within a ~50 km radius over two years. The geographic position of all sampling locations is shown in Figure 1 and characteristics of all sampled carrot populations are provided in Supplementary Table S1. Plants were sampled independently of the presence of viral symptoms but plants with obvious fungal infestation or necrosis were excluded. Leaf samples were stored desiccated over anhydrous $CaCl_2$ (Sigma-Aldrich, Saint-Quentin-Fallavier, France) until used. For each population, a pool corresponding to 50 plants was assembled by pooling 0.1g of dried leaves per plant and grinding with mortar and pestle until a fine powder was obtained and stored for a short term (1-7 days) in a 10ml Falcon tube at room temperature until further use. Between each grinding step, bench and equipment surfaces were cleaned with 10% sodium hypochlorite solution followed by 70% ethanol.

Double-stranded RNAs purification, Illumina sequencing

Double-stranded RNAs (dsRNAs) were purified from each pool of carrot leaves by two rounds of CC41 cellulose chromatography and nuclease treatment as described (Marais *et al.*, 2018). In parallel, a negative extraction control was similarly prepared using only buffer. Purified dsRNAs were converted to cDNA and randomly amplified while simultaneously adding MID tags by RT-PCR (François *et al.*, 2018; Marais *et al.*, 2018). Amplification products were purified using the MinElute PCR Purification Kit (Qiagen SAS France, Courtaboeuf, France) and their concentration determined spectrophotometrically. Equal amounts of amplification products from each pooled carrot population were prepared as independent sequencing libraries.

In the case of two populations sampled in 2019 (one wild and one cultivated corresponding to populations 19_7 and 19_9, Supplementary Table S1), dsRNAs were also independently extracted from leaves of the 50 individual plants that had contributed to the pool (75 mg per plant). Following random amplification and MID-tagging as above, equal amounts of the

amplification products were used to prepare two separate multiplexed sequencing libraries, corresponding respectively to 50 wild and 50 cultivated carrot plants. All prepared libraries were sequenced in multiplexed format (2×150 bp) on an Illumina NovaSeq 6000 system at the GetPlaGe platform (GenoToul INRAE Toulouse, France).

Bioinformatic analyses: demultiplexing and quality trimming, *de novo* assembly, contigs annotation and reads mapping to viral reference genomes

Raw sequencing reads were imported into CLC Genomics Workbench (CLC-GW), version 21.0.3, demultiplexed and trimmed on quality and length using default settings and a minimum read length of 60 nucleotides (nt). Contigs were subsequently assembled *de novo*, using a minimal length of 250 nt. Viral contigs were identified by a comparison against the non-redundant sequence (nr) database in GenBank using BlastX (Altschul *et al.*, 1990). Consensus virus genome sequences were reconstructed using *de novo* assembled contigs and, if needed, through extension by repeated mapping of remaining reads. Novel virus species were identified through calculating percent pairwise divergence with reference sequences using p-distance analysis as implemented in MEGA7 (Kumar *et al.*, 2018) and following International Committee on Taxonomy of Viruses (ICTV) species molecular demarcation criteria for each virus genus or family.

In order to compare the viromes of different carrot populations, sequencing depth for each library was first normalized by randomly subsampling reads to a common threshold using the read sampling tool in CLC-GW. Reads from normalized datasets were then mapped against a set of reference viral genomic sequences composed of reference sequences from GenBank (for known viruses) and of assembled genomic scaffolds for novel agents. Stringent mapping parameters were used (nt identity and length fraction both of 90%) in an effort to limit reads cross-mapping between closely related viruses. The mapped reads were then manually checked to ensure the identity of the detected virus. In order to address the well-known HTS inter-sample

crosstalk issue, two approaches were used to estimate it. The first one considered the viral reads identified in the blank control for each run. The second one used data from individually analyzed plants for which both RT-PCR and HTS data were available, so that plants that tested negative in RT-PCR but showed a few reads for a virus in the corresponding dataset could be used to calculate an average number of background reads per million sequencing reads. With this strategy, a separate detection threshold could be calculated for those viruses that were screened by RT-PCR. For viruses not tested by RT-PCR, we used as a threshold the average of the threshold values determined for the RT-PCR-tested viruses. Finally, the thresholds obtained for each virus using either the blank controls or the RT-PCR testing were compared and the highest value retained to establish the final HTS positive detection threshold for each virus which was used to clean the reads mapping table for each sample/virus combination.

Analysis of genomic sequences, sequence comparison and phylogenetic analyses

Multiple alignments of reconstructed genomes for known and new viruses identified in the carrot virome as well as of RT-PCR viral amplicon sequences from individual plants and of reference sequences retrieved from GenBank were prepared using ClustalW as implemented in MEGA 7 (Kumar *et al.*, 2018). Neighbour-joining trees were inferred from alignments of whole genomes or of partial nt genomic sequences using strict nt or amino acid (aa) identities calculated using MEGA 7. The significance of branches was estimated with 1,000 bootstrap replicates. Genetic differentiation between populations of viral isolates from different carrot populations was evaluated using the Fixation index (Fst) as implemented in the DnaSP6 software (Rozas *et al.* 2017). The Fst value calculates the genetic differences in terms of variances in the frequencies of the polymorphic sites between the tested viral populations. Fst values can range from 0 to 1, where values <0.18 indicate undifferentiated populations, values between 0.18 to 0.30 moderately differentiated and > 0.3 differentiated populations.

Total RNA extraction, RT-PCR detection and characterization of selected viruses

Total RNAs were individually extracted from individual plants from eight carrot populations (two cultivated, one off-type and five wild populations) using the Nucleospin 96 RNA extraction kit (Macherey-Nagel, Hœrdt, France) following the manufacturer's instructions with some modifications. Briefly 10 mg of dried plant material were transferred to a 2ml tube and one metal bead added. Samples were frozen at -80°C for at least 1h before proceeding. Frozen samples were ground using a horizontal shaker (Retsch MM301, Haan, Germany) for 1-2 min at a frequency of 30 Hz, until a fine powder was obtained. Cell lysis was performed by adding 500 µl of RAP lysis buffer with 1% ß-mercaptoethanol and 1% polyvinylpyrrolidone 40 (Sigma Aldrich). Samples were gently mixed on a vertical shaker (Heidolph Unimax 2010, Germany) at 160 rpm for 5-10 mn and subsequently centrifuged for 10 mn at 10,000 g. To reduce viscosity, 400 µl of the supernatant was transferred onto a NucleoSpin® Filter plate placed on a MN Square-well Block and centrifuged for 3 mn at 4,000 g. Binding conditions were adjusted by the addition of an equal volume of 70% ethanol to each sample. Upon completion of the RNA-binding step, membranes were desalted by adding 500 µl Membrane Desalting Buffer. Subsequent to the DNase digestion step, total RNA was purified by a wash with 500 µl RA-2 buffer and two further washes with 500 µl RA-3 buffer. RNAs were finally eluted in 96 well PCR plates by adding 50 µl RNase-free water to each sample, incubating for 2 mn and centrifuging at 4,000 g for 3-5 mn.

Specific two-step RT-PCR assays (Marais *et al.* 2011) were developed for selected viruses of the family *Solemoviridae, Tombusviridae* and for satellite RNAs. Primers were designed using multiple alignments generated from contigs of the corresponding viruses obtained from virome data and of related viral references from GenBank. Primers were designed manually, targeting conserved sequences for the virus of interest that would allow its discrimination from related viruses, thus ensuring both inclusiveness (ability to detect all isolates of a virus) and specificity

(ability to only amplify isolates of the target virus). Details on these primers and on the developed RT-PCR assays are shown in Supplementary Table S2. PCR cycling condition were as follows: 40 cycles of each at 95°C for 30s, virus-specific annealing temperature (see Supplementary Table S2) for 30s and 72°C for 30s.

Total RNAs from individual plants were analysed by RT-PCR in 96 well PCR plates and amplicons visualized following 1.2 % non-denaturing agarose gel electrophoresis. To confirm the specificity of the PCR assays and for subsequent sequence analysis of viral diversity, up to 39 positive amplicons per virus and population were directly Sanger sequenced (Eurofins Genomics, Germany).

Statistical analyses of virome fingerprints and virus associations: hierarchical clustering, bipartite network analysis and search for non-random virus co-occurrence

Statistical analyses were performed using R software 4.2.1. Virome structure was analysed in terms of virome richness as the number of species detected in different carrot populations. After verification of the distribution normality and variance homogeneity, a two-sided t-test was used (confidence level 95%). To analyse differences in the nt diversity between virus populations of different aphid transmitted viruses identified in wild and cultivated carrots a distribution-free two sided Mann–Whitney–Wilcoxon test was performed.

To further analyse virome fingerprints of different carrot populations, a hierarchical clustering was performed using the R package ‘stats’. Briefly, after z-score based scaling of mapped read numbers, non-parametric spearman rank correlation was used to generate a dissimilarity matrix. Complete linkage was used as agglomerative method. The clustering was visualized with the heatmap.2 function implemented in the “ggplot” package. To confirm the hierarchical clustering and to gain a deeper insight into the virus community structure, bipartite networks were created using the “bipartite” package in R. The network was designed with one set of nodes representing different virus taxa and the second set of nodes representing the host carrot

populations. An incidence matrix was derived from the table of mapped reads (used for the hierarchical clustering) and converted into a binary matrix indicating the presence (1, samples with ≥1 read above the positivity threshold determined as explained above) or absence (0) of the virus in each carrot population. Non-random virus-virus co-occurrences were calculated using presence/absence data from the RT-PCR-screening of aphid transmitted viruses in individual plants from five wild carrot populations. A probabilistic and distribution free model based on combinatorics implemented in the R package 'cooccur' (Griffith *et al.*, 2016) was used to calculate non-random virus associations. The algorithm compares the probabilities of observed frequencies of co-occurrence being significantly higher, lower or equal to the expected frequencies by random and then, in each case, deduces positive, negative or random associations between pairs of viruses (Veech, 2013).

RESULTS

Overview of viruses detected in cultivated and wild carrots

A total of 31 carrot populations including 11 cultivated, seven off-type and 13 wild populations sampled over two years in the Gironde and Landes departements (France) were analysed using a dsRNA-based metagenomic approach. This was part of a larger carrot virome study in France and Spain and the known and novel viruses thus identified will be described in detail elsewhere (Schönegger *et al.*, manuscript in preparation). Sequencing libraries contained on average 41% of viral reads (range 1% to 80%). The analysis of the Gironde and Landes carrots revealed a very rich virome of 36 viruses including unclassified satellite RNAs. Half of the identified viruses (18) are already known agents assigned to seven different genera in seven different families (*Solemoviridae, Tombusviridae, Betaflexiviridae, Secoviridae, Closteroviridae, Bromoviridae,* and *Partitiviridae*) or represent unclassified *Riboviridae* or subviral agents (Table 1). For these analyses, CMoVsatRNA and CMoMVsatRNA were considered as representing a single species due to their high percentage of nucleotide identity (93%) and to

the ensuing difficulty to avoid reads cross-mapping between them. The remaining half (18 viruses) are new viruses of the family *Solemoviridae, Tombusviridae, Closteroviridae, Betaflexiviridae, Partitiviridae, Totiviridae* and *Amalgaviridae* as well as two putative novel *Riboviria* and four subviral agents, two of which are circular hammerhead satellites recently described by Babalola *et al.* (2022) and one a tombusvirus-like associated RNA (tlaRNA) recently described by and Schönegger *et al.* (2022) (Table 1). A detailed description of these new agents will be presented elsewhere, together with other new carrot viruses identified in the virome of carrots sampled in Spain or in other regions of France (Schönegger et al., manuscript in preparation; Candresse et al., manuscript in preparation). A high temporal stability of the virome between sampling years was observed with 81% of viruses detected in both sampling years. The seven viruses responsible for the between years differences are rare viruses that were only detected in a single carrot population in one of the two years.

Differentiation of the virome between wild and cultivated carrot populations

The virome richness, defined as the number of virus species detected in each carrot population type over two years showed a significant difference between cultivated and wild carrots (p-value = 0.004) with a higher virome richness in wild carrot populations (15.5±2.4 viruses compared to 12.2±2.6). Off-type carrot populations had an intermediate richness (14±1.8 viruses), so that the differences with cultivated or wild populations were not statistically supported. A hierarchical clustering (HC) was performed using a weighted and scaled matrix based on the number of mapped reads against reference viral sequences. The analysis shows a strong separation in the clustering between wild and field carrot populations (Figure 2). Off-type populations cluster together with cultivated populations (red cluster in Figure 2). The off-type populations are morphologically and phenologically distinct from cultivated carrots and show some wild carrot traits suggesting they are either wild carrots or hybrids between wild and cultivated carrots that contaminated the commercial carrot seed lots. Nevertheless, they are

grouped together with the cultivated carrots, reflecting similarities in their virome structure. No clustering according to time or location was observed.

More than half of the 36 identified viruses were detected in wild carrots (24 viruses, 66%), and four (16.7%) out of these were exclusively found in wild populations (Figure 3). These viruses comprise a known polerovirus, wild carrot red leaf virus (WCtRLV), a novel polerovirus, carrot polerovirus 1 (CaPV1, Schönegger et al., manuscript in preparation), a new tlaRNA, carrot red leaf virus-associated RNA 2 (CtRLVaRNA2, Schönegger *et al.*, 2022) and an Ilarvirus reported here for the first time in carrots, solanum nigrum ilarvirus 1 (SnIV1, Ma *et al.*, 2020). A similar pattern is seen in carrot populations sampled in cultivated fields with 75% (27) and 72.2% (26) viruses identified in cultivated and off-type populations, respectively. Remarkably, cultivated and off-type carrots have four viruses (11.1%) in common that are not found in wild carrots, when the corresponding value for shared viruses between cultivated and wild carrots or cultivated and off-type carrots is only two in each case (Figure 3). The viruses shared by cultivated and off-type carrots belong to the genera *Closterovirus* and *Chordovirus* and include also a tlaRNA, beet western yellows-associated RNA (BWYVaRNA) and a new unclassified virus showing distant homologies to members of the *Benyviridae* family, carrot-associated RNA virus 1 (CaRNAV1, Fox *et al.*, 2022). No member of the *Closterovirus* and *Chordovirus* genera was found in wild carrots. About half of the viruses (47.24%, 17 viruses) are shared among all three population types (Figure 3). These viruses include mostly aphid-transmitted viruses such as members of the families *Solemoviridae* and *Tombusviridae* and satellite RNAs that are associated with them, but also one member of the *Secoviridae* family, carrot torradovirus 1 (CaTV1). Interestingly, all three population types also share seed-borne viruses of the family *Partitiviridae*, including a divergent isolate of dill cryptic virus 2 (DCV2) with respectively 83% and 82% nt identity over genomic RNAs 1 and 2 with the DCV2 reference isolate (NC_021147 and NC_021148). Another shared persistent virus is carrot cryptic virus

(CaCV), which however was detected in 82% and 86% of the cultivated and off-type populations, respectively, but was found in only one out of the 13 wild carrot populations (8%). A detailed listing of the viruses in different parts of the Venn diagram presented in Figure 3, with the corresponding number of infected populations in both years is given in Supplementary Table S3.

Bipartite network analysis confirms contrasted viromes between wild and field carrot populations

As a complementary virome fingerprint analysis, we used a bipartite network analysis based on an unweighted binary incidence matrix derived from the mapped reads matrix. In the network, one type of nodes corresponds to viruses and the other type of nodes to carrot populations (Figure 4). The network shows a division in the graph: in the right part, a concentration of nodes corresponding to field carrot populations (cultivated in red and off-type in orange) and in the left part, nodes corresponding to wild populations (in green). This pattern confirms the separation seen in the hierarchical clustering, even though it is less pronounced, which can be seen in the overlap of a cultivated population (20_2, Supplementary Table S1) with the cluster of wild populations in the network (Figure 4), indicating greater similarity of its virome with those of wild populations. Interestingly, population 20_2 was sampled on the outer border of a cultivated carrot field and directly adjacent to wild population 20_6 that bordered this field (Supplementary Table S1). Taking a closer look at both populations, 20_2 was the only cultivated population in 2020 in which the novel viruses CaEV1 and CtFLV1 were identified. Both of them were also identified in the adjacent wild 20_6 population. To determine whether the similarity in their virome was due to increased viral exchange, facilitated by the proximity of the cultivated population to the neighboring wild population, HTS contigs assembled for viruses shared between the two populations (two new and one known umbraviruses, CaUV2, CaUV3, and CMoV respectively) were extracted and compared in multiple alignments. CaUV2

and CMoV contigs showed 100% nt identity in overlap regions of respectively 1,063 and 300 nt, while CaUV3 contigs showed a single nucleotide polymorphism over a 1,252 nt shared region. The same viruses show a lower sequence identity when compared to isolates from populations more distant in the network. For example, the isolates of CaUV2, CaUV3 and CMoV from the population most distant to 20_2 in the network, 19_3, showed respectively 29, 67 and 7 mutations in a contigs overlap region of 1,170 nt. This observation of identical or nearly identical isolates shared by neighboring wild and cultivated populations is suggestive of the existence of viral exchanges between them.

The network groups viruses that are highly connected (identified in all or most populations) in the center of the graph as larger overlapping nodes. Among those are CtRLV, WCMoV, CMoV, CMoMV and CMoVsatRNA and CtRLVaRNA, all of which are key-members of the CMD disease complex. Other viruses which cluster also in the center of the network are preferentially linked either to wild populations (e.g. CaUV2 and CaUV3, Supplementary Table S3) or to cultivated/off type populations (e.g. CaTV1, CaUV1 or ALVEaRNA, Supplementary Table S3). Viruses that are less connected (rare viruses) and/or those being specifically identified in one of the population types are positioned at the periphery of the network with CaPV1 or WCtRLV on the periphery of the wild nodes cluster and members of the genera *Closterovirus* and *Chordovirus* on the periphery of the cultivated cluster. The network effectively allows to visualize the different viruses present and shared (Figure 3, Supplementary Table S3) and confirms the different virome fingerprints between wild and field carrots seen in the hierarchical clustering (Figure 2). The network analysis also provides information at a qualitative level (presence/absence) that complements the quantitative level (number of reads) of the hierarchical clustering. The network analysis shows no grouping in terms of year of sampling or location, except for the case of the cultivated 20_2 population discussed above.

Population genetics of aphid-transmitted viruses

To gain insights into the prevalence and diversity of aphid-transmitted viruses, individual plants from eight populations including two cultivated, one off-type and five wild populations (19_1, 19_2, 19_7, 19_6, 19_9, 19_12, 19_13, 19_14, Supplementary Table S1) were screened by RT-PCR for the presence of 16 viruses including CtRLV, CaPV1, CaPV2, CMoV, CMoMV, WCMoV, CaUV1, CaUV2, CaUV3, CtRLVaRNA, CtRLVaRNA2, ALVEaRNA, BWYVaRNA, CMoVsatRNA and CtMoVsatRNA2. The analysis revealed a significant difference in the prevalence of known and novel viruses between the cultivated and wild carrot populations (Table 2).

With the exception of BWYVaRNA, which was only detected in a single plant of a single population, all other tested known viruses show a similar pattern with generally higher infection rates in field populations than in wild ones. The polerovirus CtRLV was thus detected in 100% of cultivated or off-type plants tested, as compared to a 14-42% prevalence in wild populations. Corresponding values for the CtMoV, CtMoMV, and WCMoV umbraviruses are respectively 80-100%, 10-70% and 20-88% for field populations, to be compared with 16-64%, 9-30% and 14-36% for wild populations, respectively (Table 2). The same pattern is again seen for associated RNAs and satellites (CtRLVaRNA, ALVEaRNA and CMoVsatRNA) with respectively 86-100%, 29-98% and 33-96% prevalence in field populations to be compared with respectively 9-54%, 0-10% and 16-40% prevalence in wild populations (Table 2).

An opposite pattern is observed for most novel viruses, which generally show a higher prevalence in wild populations than in field ones from which the novel viruses are usually absent or show only sporadic infection of one to two plants per population (Table 2). One notable exception to this trend is CaUV1 which showed a 10-12% prevalence in cultivated carrots, 3% in the tested off-type population but was not detected in wild carrots (Table 2).

RT-PCR results from individually analyzed plants comprising five wild populations (19_7, 19_6, 19_9, 19_12, 19_13, 19_14, Supplementary Table S1) were used to construct a binary matrix (1=present, 0=absent) of the 16 aphid-transmitted viruses tested that comprise the following three groups: 1) helper viruses of the family *Solemoviridae*, 2) dependent viruses of the family *Tombusviridae* and 3) dependent subviral agents (satellites and tlaRNAs). A co-occurence analysis revealed 25 positive associations between pairs of agents (Figure 5). Interestingly, most positive associations were observed between dependent agents with the highest number between members of the *Umbravirus* genus (50%, 12/25) and between associated subviral agents and various umbraviruses (20%, 5/25) (Figure 5). Only 16% (6/25) of positive associations were found between *Solemoviridae* helper viruses and dependent *Tombusviridae* members (Figure 5). Two positive associations (8%) were found between helper viruses and associated satellites which in both cases are novel: CaPV1 and CaEV1 (helper viruses) were positively associated with CtRLVaRNA2 and CMoVsatRNA2 (dependent agents), respectively. A single positive association between two helper viruses, CaPV1 and CaEV1 was also observed. In general, the observed positive associations suggest a low level of specificity; for example, CtRLVaRNA was positively associated with three umbraviruses, CMoV, CMoMV and WCMoV. No statistically supported negative association was identified (Figure 5).

RT-PCR amplicons and HTS contigs obtained for aphid-transmitted agents from individually analysed plants were used to generate multiple alignments for each virus consisting of up to 66 sequences and additional closely related sequences from GenBank.

These alignments were used to reconstruct neighbour joining phylogenetic trees and to calculate the degree of genetic differentiation between viral populations from different types of carrots by computing Fst values.

Populations of novel aphid-transmitted viruses (including CaUV1, CaUV2, CaUV3, CaPV1, CaPV2, CaEV1, CtRLVaRNA2 and CMoVsat2) being mainly present in wild carrots with an overall low prevalence (Table 2), did not show genetic differentiation according to Fst calculations or phylogenetic analyses (Supplementary Figure S2)

The same is true for already known aphid-transmitted viruses, with the exception of WCMoV and CtRLV which showed evidence of genetic differentiation between carrot population types with Fst values of 0.23 and 0.36, respectively. In parallel, besides a main cluster grouping isolates from different carrot populations, the phylogenetic analysis of WCMoV and CtRLV isolates indeed reveals for both viruses the existence of a separate cluster of isolates specific of field (WCMoV) or wild (CtRLV) carrots, respectively (Figures 6A and 6B).

Nucleotide diversity could be compared between wild and field populations using the multiple sequence alignments used for the phylogenetic analyses. For three of them, CtRLV, CMoV and CtMoMV, intraspecific diversity was significantly higher in wild carrots than in cultivated ones (Table 3). Conversely, WCMoV showed a higher diversity in cultivated carrots than in wild ones but the difference was not statistically significant (Table 3). As previously, viral populations from off-type carrots tended to show an intermediate behaviour. Only in the case of CtRLV the differences with both cultivated and wild carrot isolates was statistically significant (Table 3). Differences with wild carrot isolates were also significant for CMoV and WCMoV while that with cultivated carrot isolates was significant for CMoMV (Table 3). Since the diversity patterns observed for CtRLV and WCMoV might be a consequence of the existence of host-specific clusters of isolates (Figures 6A-B), the diversity was calculated again, after excluding the isolates from host-specific clusters. In the case of CtRLV, diversity of wild carrots isolates was still significantly higher (3.53%± 0.60%, p-value < 0.000) than that of cultivated carrots isolates (Table 3). In the case of WCMoV, removal of isolates from the field carrot-specific cluster resulted in a lower diversity of 2.11%± 0.24%, which is even lower than

the values observed for off-type or wild carrots (Table 3) and is statistically different from both with a p-value of 0.007 and 0.000, respectively.

DISCUSSION

In the present study, we compared the virome of cultivated (*Daucus carota* ssp. *sativus*) and wild (*D. carota* ssp. *carota*) carrot populations using a dsRNA-based virome characterization approach. Most metagenomic studies to date have compared plant viromes between wild and cultivated plants belonging to different species, genera or even families (Bernardo *et al.*, 2018, Ma *et al.* 2020, Ma *et al.* 2021, Maachi *et al.* 2022). Here, carrots were used as a pathosystem in which the genetic barriers to virus movement are expected to be the lowest given that the hosts analysed belong to the same botanical species and that carrot domestication is relatively recent. Virome data at host population level were integrated with data at the level of individual plants by combining HTS-based virome characterization with RT-PCR amplicon sequencing. The analysis at these two different levels of resolution, from populations to individual plants, allowed to contrast specific virome fingerprints or prevalence rates of individual viruses and, for selected viruses, to analyse viral population genetics and viral movement at the agroecological interface.

The dsRNA-based metagenomic approach used is able to capture all viruses with RNA genomes and to provide a powerful and inclusive representation of complex plant-associated viromes (Marais *et al.* 2018; Ma *et al*., 2019, 2020 2021). From 31 carrot populations, we identified a rich virome of 36 viruses, of which 18 are novel (Table 1). This virome is dominated by (+) ssRNA viruses, most of them being aphid transmitted, followed by dsRNA viruses. It also includes a substantial part of subviral agents including tlaRNAs and single-stranded, linear or circular satellite RNAs that are associated with aphid-transmitted viruses. This virome

appears to be relatively stable over time, since 81 % of the viruses were identified in both sampling years.

In order to capture a large fraction of the virome, including viruses present at a low prevalence in the sampled carrot populations, we selected to analyse pools of 50 plants, similar to the strategy successfully used by Ma *et al.* (2020). HTS-sequencing of individual plants of a wild population (19_9) and a cultivated population (19_7, Supplementary Table S1) showed that 94% of the viruses identified by single-plant analyses were identified by the analysis of the pool containing these plants, giving confidence in the representativeness and completeness of the virome descriptions achieved. A frequent problem in viral metagenomic studies is reads cross-talk or "bleeding" between samples, which can result from experimental contaminations but also from low level index hoping during Illumina sequencing (Illumina, 2017; van der Valk, 2019). In order to limit this problem, a double strategy was used here to develop a reliable positivity threshold for mapped reads of each virus, conservatively taking into account both read numbers in negative, buffer only controls, and comparison of HTS read numbers with RT-PCR results from the testing of individual plants. This ensures a very low probability of false-positive detections that could otherwise lead to the incorrect reporting of the presence of a virus in an analysed population.

Using different strategies to analyse virome structure, including clustering and bipartite network approaches, we observed that field (i.e. cultivated plus off-type) and wild carrot populations show contrasted virome fingerprints (Figures 2 and 4). This is despite the fact that the two approaches rely on different types of data since hierarchical clustering takes into account the matrix of mapped reads against reference sequences (and therefore mapped reads numbers), while the network analysis only uses presence/absence data. This convergence of the two types of analyses gives strong credibility to the obtained results. Off-type carrot populations clustered together with cultivated carrot populations in both analyses (Figures 2 and 4). Off-type

populations were sampled in cultivated carrot fields but are morphologically and phenologically very different from cultivated carrots, showing early bolting and thin taproots typical of wild carrots. These tap roots were frequently white but in some instances were orangy, prompting the hypothesis that off-type plants might result from hybridization between wild and cultivated carrots in seed production fields. When analysed collectively, cultivated and off-type carrot populations share a slightly larger number of viruses (21) as compared to wild carrots with either cultivated or off-type populations (19) (Figure 3). The viruses they uniquely share are a novel, unclassified virus, CaRNAV1 (Fox et al., 2022), showing distant homologies to *Benyviridae* family members and viruses from the genus *Closterovirus* and *Chordovirus* (Figure 3, Supplementary Table S3). These families seem to be associated with the cultivated compartment as their members were uniquely detected in field carrot populations. This observation is in line with the findings of Bernardo *et al.* (2018) suggesting that some viral families can be preferentially associated with cultivated plants. However, members of the *Closteroviridae* family were not identified among those associated with agriculture in that study. It must also be noted that viruses found exclusively in field populations in this study are in most cases rare viruses that were detected in only 2-4 of the 18 populations analysed (Supplementary Table S3), so that their specificity could reflect a limitation in the number of wild populations sampled. A notable exception is CaRNAV1, which was detected in all 18 field populations but was absent from all 13 wild populations (Supplementary Table S3), even when sampled at the border of a field. Given its very recent discovery (Fox *et al.*, 2022), there is very little information available on CaRNAV1 and it is today unclear whether it infects carrots or might infect another organism associated with carrots, such as a fungus.

Most agents found specific for wild populations were detected in most or all tested populations (Supplementary Table S3), with the exception of the ilarvirus SnIV1 which was only detected in 3/13 populations. Remarkably, the other three agents specifically associated with wild carrot

populations, two poleroviruses (WCtRLV and CaPV1) and a tlaRNA (CtRLVaRNA-2) are expected to be transmitted by aphids.

The differences in virome structure were paralleled by a statistically significant difference in virome richness at the viral species level between cultivated and wild populations, with higher virome richness in wild populations (15.5±2.4) compared to cultivated populations (12.2 ±2.6). This finding is in line with previous studies showing that wild plants are frequently infected with viruses and commonly harbour new viruses (Melcher *et al.* 2008, Muthukumar *et al.* 2009, Roossinck *et al.*, 2010, Thapa *et al.*, 2015, Bernardo *et al.*, 2018, Ma *et al.* 2020, Rivarez *et al.*, 2022). However, it contrasts with the studies of Bernardo *et al.* (2018) and Ma *et al.* (2020) which revealed a higher virus richness in cultivated than in wild plants. However, it is difficult to compare these studies with the present results since the wild and cultivated species used were not as botanically close as here and since these studies either analyzed virome richness at family level (Bernardo *et al.;* 2018) or used OTUs as a proxy to species (Ma *et al.;* 2020).

Besides differences in virome composition, strong differences in the prevalence of aphid-transmitted viruses between field and wild populations were observed. The prevalence of known aphid-transmitted viruses was much higher in field carrots than in wild carrots for CtRLV, umbraviruses and associated subviral agents (Table 2). These results are in line with other studies at landscape level showing higher virus prevalence in managed or cultivated areas compared to less managed or wild areas (Pagán *et al.* 2012; Bernardo *et al.* 2018). The degree of human management has also been shown to impact aphids abundance and diversity (Al Hassan *et al.*, 2013). Claflin *et al.* (2016) observed that crop dominated landscapes show a higher prevalence of the vector-transmitted potato virus Y and that aphids abundance is positively correlated with virus prevalence. The high density of host plants in cropping systems likely leads to an increased frequency of transmission and therefore to an increase in viral prevalence for epidemic viruses.

These viruses are transmitted by the willow carrot aphid, *Cavariella aegopodii*, an heteroecious species that alternates between primary, overwintering hosts (*Salix spp*.) and secondary hosts in the *Apiaceae* family (Dunn, 1965, Williams and Dixon, 2007). On the secondary hosts, several wingless generations are produced parthenogenetically and damaging dense aphid populations can develop. Migration of aphids from willows to carrots peaks in June and declines in late summer when winged, sexual forms are produced that migrate back to their primary hosts in autumn to mate and lay their eggs (Dunn, 1965; Williams and Dixon, 2007). The carrots populations analyzed individually in this study were sampled at the end of July, providing time for the spring flights to introduce virus inoculum in carrot fields and for further local spread by wingless forms in the dense field carrot populations, thus likely explaining the high prevalence levels observed.

As shown in previous studies, wild plants are an important source of unknown viral diversity (Cooper and Jones, 2006, Roossinck, 2011; Roossinck, 2012) and a high proportion of novel viruses or variants were frequently found in unmanaged systems (Bernardo *et al.* 2018; Ma *et al.*, 2020; Maachi *et al.* 2022, Rivarez *et al.* 2022). The finding of numerous new viruses in wild carrots is therefore not surprising. However, with the exception of CaUV1, the prevalence pattern of novel viruses belonging to aphid-transmitted genera was opposite to that of known viruses of the same genera, generally showing a higher prevalence in wild populations (4-20%) compared to field populations in which they were generally absent or infected only sporadically one or two plants per sampled population (Table 2). This raises the question of the origin of this difference in behavior. Based on single plant screening of five wild populations (the field populations data were not informative given the very high or very low prevalence rates), we analysed virus co-occurrence between known and novel members of the *Solemoviridae* and *Tombusviridae* families and associated subviral agents (Figure 5), as they are known to be potentially involved in assistance and complementation interactions. Indeed, the genome of

umbraviruses encodes an RdRp as well as two other proteins involved in cell-to-cell and long distance movement but lack a capsid protein necessary for the formation of transmissible virus particles (Syller, 2003). Consequently, umbraviruses depend on encapsidation by the capsid protein of poleroviruses or enamoviruses for their aphid transmission, while the movement of helper virus(es) may be facilitated by the movement proteins of the umbravirus (Syller, 2003 and Yoshida, 2020). The analysis revealed 25 positive associations, with more than two thirds of them between dependent agents (umbraviruses and associated subviral agents). Surprisingly, only about 25% of positive association were between helper viruses and dependent agents (Figure 5). This was an unexpected result considering the strict dependence of umbraviruses and associated subviral agents on polerovirus/enamovirus assistance for encapsidation and aphid transmission. Two reasons could explain this low level of detection of the most expected interactions. The first is that promiscuity of dependent agents, with an ability to be assisted by different helper viruses would greatly affect the ability to detect statistically significant co-occurences. The second would imply that although assisted, a dependent agent would not necessarily be simultaneously transmitted with its helper virus. Indeed, we observed that out of the 80 individually tested plants that showed mixed infections of umbraviruses with subviral agents, 33 (41%) did not show evidence for a concomitant infection by a polerovirus or enamovirus helper virus. This surprising finding parallels previous results on tlaRNAs showing coinfection of umbraviruses and tlaRNAs in the absence of helper viruses as demonstrated by HTS analysis of individual carrot plants (Schoenegger *et al.* 2022). However, a positive association could be observed between the key players of the CMD complex, the polerovirus CtRLV and the umbravirus CMoV. Yoshida *et al.* (2020) observed that a coinfection with CMoV and CtRLV can increase disease severity and RNA accumulation of CtRLV, while the RNA accumulation of CtRLV was abolished in the presence of CtRLVaRNA. The authors also found that CtRLVaRNA only occurred in coinfections with CMoV but never with CtRLV

alone. It can therefore be hypothesized that dependent viruses (umbraviruses or associated subviral agents) can be transmitted by vectors without concomitant transmission of the helper virus, in particular if the coinfecting agents have a negative impact on the helper virus accumulation (Yoshida, 2020). Umbraviruses could further assist subviral agents for movement within the plant but such a scenario would represent a dead-end infection without the encapsidation by the concomitant helper virus. Overall, the co-occurrence analysis suggests that there is little specificity in the associations between carrot dependent and helper viruses (Figure 5A), indicating a significant level of promiscuity in this system of assistance and complementation. In particular, when considering the novel agents preferentially associated with wild carrots (CaPV1, CaPV2, CaEV1, CaUV2, CaUV3, CtRLVaRNA-2, CMoVsatRNA-2), positive associations were detected both within this group of agents but also with known agents showing preferentially higher prevalence levels in cultivated carrots (Figure 5), weakening the hypothesis that a different aphid specificity might be responsible for the preferential association with wild carrots.

To gain deeper insight into the movement of aphid-transmitted viruses between wild and cultivated carrots, we compared virus populations infecting the three different carrot population types through RT-PCR amplicon sequencing and HTS analysis of single plants. In the case of known aphid-transmitted viruses, a sufficient number of isolates were analyzed to compare nucleotide diversity between viral populations from wild, off-type and cultivated carrots. Three of these viruses, CtRLV, CtMoV and CMoMV show a higher intraspecific diversity in wild carrots compared to cultivated carrots (p-values <0.00). In the case of CtRLV, this was paralleled by the identification of a phylogenetic cluster of isolates specifically associated with wild carrots (Figure 6B). For WCMoV, a reverse trend was observed but was not statistically significant (Table 3), and a cluster of isolates specifically associated with field carrots was identified (Figure 6A). Taken together, these results suggest the existence of barriers to the

spread of some isolates of CtRLV and WCMoV between the wild and cultivated compartment. On the other hand, other isolates of these two viruses seem to be able to freely circulate between wild and field carrots, as evidenced by the lack of host population-based separation of isolates in the main phylogenetic clusters (Figure 6A-B). It is however not possible to decide on the directionality of the viral fluxes, from field to wild (spill-over), or from wild to field (spill-back from a wild reservoir). These results parallel those of Maachi *et al.* (2022), who identified a specific lineage of curcubit aphid borne yellows virus (CABYV) in the wild plant *Ecballium elaterium* that was genetically distinct from CABYV isolates in melon crops and those of Ma *et al.* (2020) for potato virus Y in the tomato/nightshade pathosystem.

Taken together, the results reported here demonstrate surprisingly different virome fingerprints between wild and field carrots. This is in part due to differences in persistent viruses but differences were also observed for acute viruses, most of which are expected to be aphid-borne. Despite their different genetic makeup, off-type carrots had a virome that was closely related to that of cultivated carrots growing side by side with them, suggesting a major impact of local conditions, including those created by agriculture with dense populations of actively growing plants, in shaping virome richness, composition and individual virus prevalence. There is evidence that the host genotype and the environment are important drivers in shaping virus communities (Sallinen *et al.* 2021) and that changes in the nutrient supply (e.g. phosphorus or nitrogen) can alter prevalence and interactions among coinfecting viruses (Lacroix *et al.* 2014). While evidence was obtained for the existence of viral fluxes between the wild and cultivated compartments for most aphid-transmitted viruses, population genetics analyses indicated the existence of barriers limiting or altogether preventing these fluxes for some viruses or for some phylogenetic groups within other species. The bases for these barriers, which might reflect differential susceptibility of carrot populations, vector preferences or yet other sources remain to be identified.

Acknowledgements: The authors wish to thank the INRAE GenoToul Platform (Toulouse, France) for Illumina sequencing.

Funding: This study was funded by the European Union through a Horizon 2020 Marie Skłodowska-Curie Actions Innovative Training Network (H2020 MSCA- 60 ITN) project "INEXTVIR" (GA 813542).

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

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

REFERENCES

Adams IP, Skelton A, Macarthur R, Hodges T, Hinds H, Flint L, Nath PD, Boonham N, Fox A (2014) Carrot yellow leaf virus is associated with carrot internal necrosis. PLOS One 9, e109125. https://doi.org/10.1371/journal.pone.0109125

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990) Basic local alignment search tool. J Mol Biol. 215, 403-410. https://doi: 10.1016/S0022-2836(05)80360-2

Anon (2018) La production fruitière et légumière, un ancrage régional. Agreste Nouvelle Aquitaine, 7, 1-4. https://draaf.nouvelle-aquitaine.agriculture.gouv.fr/IMG/pdf/AgresteNA_FI_no07_nov2018_FiliereFruitslegumesSalon_cle4fea83.pdf

Babalola BM, Schönegger D, Faure C, Marais A, Fraile A, Garcia-Arenal F, Candresse T (2022) Identification of two novel putative satellite RNAs with hammerhead structures in the virome of French and Spanish carrot samples. Arch. Virol. 167, 2287-2292. https://doi.org/10.1007/s00705-022-05538-z

Bernardo P, Charles-Dominique T, Barakat M, Ortet P, Fernandez E, Filloux D, Hartnady P, Rebelo TA, Cousins SR, Mesleard F, Cohez D, Yavercovski N, Varsani A, Harkins GW, Peterschmitt M, Malmstrom CM, Martin DP, Roumagnac P (2018). Geometagenomics illuminates the impact of agriculture on the distribution and prevalence of plant viruses at the ecosystem scale. ISME J. 12, 173-184. https://doi.org/10.1038/ismej.2017.155

Brunt AA, Crabtree K, Dallwitz MJ, Gibbs AJ, Watson L, Zurcher EJ (1996). Plant Viruses Online: Descriptions and Lists from the VIDE Database. Version: 20th August 1996. http://biology.anu.edu.au/Groups/MES/vide/

Burdon JJ, Chilvers GA (1982) Host density as a factor in plant disease ecology. Ann. Rev. Phytopath. 20, 143-166. https://doi.org/10.1146/annurev.py.20.090182.001043

Cooper I, Jones RAC (2006) Wild Plants and Viruses: Under-Investigated Ecosystems. Adv. Virus Res. 67, 1-47. https://doi.org/10.1016/S0065-3527(06)67001-2

Claflin SB, Jones LE, Thaler JS, Power AG (2017) Crop-dominated landscapes have higher vector-borne plant virus prevalence. J Appl Ecol, 54, 1190–1198. https://doi.org/10.1111/1365-2664.12831

Dunn JA (1965) Studies on the aphid, *Cavariella aegopodii* Scop. Ann. Appl. Biol. 56, 429–438.

Elena SF, Fraile A, García-Arenal F (2014) Evolution and emergence of plant viruses. Adv. Virus Res. 88, 161-191. https://doi.org/10.1016/B978-0-12-800098-4.00003-9

Fox A, Gibbs AJ, Fowkes AR, Pufal H, McGreig S, Jones RAC, Boonham N, Adams IP (2022) Enhanced apiaceous potyvirus phylogeny, novel viruses, and new country and host records from sequencing *Apiaceae* samples. Plants 11, 1951. https://doi.org/10.3390/plants11151951

Fraile A, McLeish MJ, Pagán I, González-Jara P, Piñero D, García-Arenal F (2017). Environmental heterogeneity and the evolution of plant-virus interactions: Viruses in wild pepper populations. Virus Res. 241, 68-76. https://doi.org/10.1016/j.virusres.2017.05.015

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. In: Viral Metagenomics: Methods and Protocols, Pantaleo V, Chiumenti M (Eds.), Humana Press, New York (USA), pp 77-95.

Griffith DM, Veech JA, Marsh CJ (2016) Cooccur: probabilistic species cooccurrence analysis in R. J. Stat. Softw. 69, 1-17. https://doi.org/10.18637/jss.v069.c02

Al Hassan D, Georgelin E, Delattre T, Burel F, Plantegenest M, Kindlmann P, Butet A (2013) Does the presence of grassy strips and landscape grain affect the spatial distribution of aphids and their carabid predators? Agric. For Entomol, 15, 24–33. https://doi.org/10.1111/j.1461-9563.2012.00587.x

Haňcinský R, Mihálik D, Mrkvová M, Candresse T, & Glasa M (2020) Plant viruses infecting Solanaceae family members in the cultivated and wild environments: A review. Plants, 9, 667.https://doi.org/10.3390/plants9050667

Illumina (2017) Effects of index misassignment on multiplexing and downstream analysis. https://www.illumina.com/content/dam/illumina-marketing/documents/products/whitepapers/index-hopping-white-paper-770-2017-004.pdf

Jones RAC (2014) Plant virus ecology and epidemiology: historical perspectives, recent progress and future prospects. Ann. Appl. Biol. 164, 320-347. https://doi.org/10.1111/aab.12123

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol*. **35,** 1547-1549. https://doi: 10.1093/molbev/msy096.

Lacroix C, Seabloom EW, Borer ET (2014) Environmental nutrient supply alters prevalence and weakens competitive interactions among coinfecting viruses. New Phytol. 204, 424-433. https://doi.org/10.1111/nph.12909

Latham LJ, Jones RAC (2004) Carrot virus Y: symptoms, losses, incidence, epidemiology and control. Virus Res. 100, 89-99. https://doi.org/10.1016/j.virusres.2003.12.017

Lefeuvre P, Martin DP, Elena SF, Shepherd DN, Roumagnac P, Varsani A (2019). Evolution and ecology of plant viruses. Nat. Rev. Microbiol. 17, 632-644. https://doi.org/10.1038/s41579-019-0232-3

Ma Y, Fort T, Marais A, Lefebvre M, Theil S, Vacher C, Candresse T (2021) Leaf-associated fungal and viral communities of wild plant populations differ between cultivated and natural ecosystems. Plant-Environ. Interact. 2, 87-99. https://doi.org10.1002/pei3.10043

Ma Y, Marais A, Lefebvre M, Theil S, Svanella-Dumas L, Faure C, Candresse T (2019) Phytovirome analysis of wild plant populations: comparison of double-stranded RNA and virion-associated nucleic acid metagenomic approaches. J Virol. 94, e01462-19. https://doi.org/10.1128/JVI.01462-19

Ma Y, Marais A, Lefebvre M, Faure C, Candresse T (2020). Metagenomic analysis of virome cross-talk between cultivated *Solanum lycopersicum* and wild *Solanum nigrum*. Virology, 540, 38-44. https://doi.org/10.1016/j.virol.2019.11.009

Maachi A, Donaire L, Hernando Y, Aranda MA (2022). Genetic differentiation and migration fluxes of viruses from melon crops and crop edge weeds. Journal of Virology, 96, e0042122. https://doi.og/10.1128/jvi.00421-22

Mahlanza, T. Pierneef RE, Makwarela L, Roberts R, van der Merwe M (2022) Metagenomic analysis for detection and discovery of plant viruses in wild Solanum spp. in South Africa. Plant Pathol. 71, 1633-1644. https://doi.org/10.1111/ppa.13585

Malmstrom CM, Alexander HM (2016). Effects of crop viruses on wild plants. Curr. Opin. Virol. 19, 30-36. https://doi.org/10.1016/j.coviro.2016.06.008

Marais A, Faure C, Bergey B, Candresse T (2018) Viral double-stranded RNAs (dsRNAs) from plants: alternative nucleic acid substrates for high-throughput sequencing. Meth. Mol. Biol. 1746, 45-53. https://doi.org/10.1007/978-1-4939-7683-6_4

Marais A, Svanella-Dumas L, Barone M, Gentit P, Faure C, Charlot G, Ragozzino A, Candresse T (2011) Development of a polyvalent RT-PCR detection assay covering the genetic diversity of cherry capillovirus A (CVA). Plant Pathol. 61, 195-204. https://doi.org/ 10.1111/j.1365-3059.2011.02488.x

Melcher U, Muthukumar V, Wiley GB, Min BE, Palmer MW, Verchot-Lubicz J, Ali A, Nelson RS, Roe BA, Thapa V, Pierce ML (2008) Evidence for novel viruses by analysis of nucleic acids in virus- like particle fractions from *Ambrosia psilostachya*. J. Virol. Meth. 152, 49-55. https://doi.org/10.1016/j.jviromet.2008.05.030

Menzel W, Maiss E, Vetten HJ (2009) Nucleotide sequence of a satellite RNA associated with carrot motley dwarf in parsley and carrot. Virus Genes, 38, 187-188. https://doi.org/10.1007/s11262-008-0302-5

Mihara T, Nishimura Y, Shimizu Y, Nishiyama H, Yoshikawa G, Uehara H, Hingamp P, Goto S, Ogata H (2016) Linking virus genomes with host taxonomy. Viruses, 8, 66. https://doi.org/0.3390/v8030066

Moury B, Fabre F, Hébrard E, Froissart R (2017). Determinants of host species range in plant viruses. J Gen Virol. 98, 862-873. https://doi: 10.1099/jgv.0.000742

Muthukumar V, Melcher U, Pierce M, Wiley GB, Roe BA, Palmer MW, Thapa V, Ali A, Ding T (2009) Non-cultivated plants of the Tallgrass Prairie Preserve of northeastern Oklahoma frequently contain virus-like sequences in particulate fractions. Virus Res. 141, 169-173. https://doi.org/10.1016/j.virusres.2008.06.016

Naseem MT, Ashfaq M, Khan AM, Kiss Z, Akhtar KP, Mansoor S (2016) Transmission of viruses associated with carrot motley dwarf by *Myzus persicae*. J. Plant Pathol. 98, 581-585. http://www.sipav.org/main/jpp/index.php/jpp/article/view/3698/2353

Pagán I, González-Jara P, Moreno-Letelier A, Rodelo-Urrego M, Fraile A, Piñero D, García-Arenal F (2012) Effect of biodiversity changes in disease risk: exploring disease emergence in a plant-virus system. PLOS Pathogens, 8, 47. https://doi.org/10.1371/journal.ppat.1002796

Praciak A (2022) *Daucus carota* (carrot). In: Invasive Species Compendium. CAB International, Wallingford (UK). https://www.cabidigitallibrary.org/doi/10.1079/cabicompendium.18018

Rivarez MPS, Pecman A, Bačnik K, Maksimović Carvalho Ferreira O, Vučurović A, Seljak G, Mehle N, Gutiérrez-Aguirre I, Ravnikar M, Kutnjak D (2022) In-depth study of tomato and weed viromes reveals undiscovered plant virus diversity in an agroecosystem. bioRxiv. https://doi.org/10.1101/2022.06.30.498278

Rong J, Lammers Y, Strasburg JL, Schidlo NS, Ariyurek Y, de Jong TJ, Klinkhamer PG, Smulders MJ, Vrieling K, (2014) New insights into domestication of carrot from root transcriptome analyses. BMC Genomics, 15, 1-15. https://doi.org/10.1186/1471-2164-15-895

Roossinck MJ (2011) The big unknown: plant virus biodiversity. Curr. Opin. Virol. 1, 63-67. https://doi.org/10.1016/j.coviro.2011.05.022

Roossinck MJ, García-Arenal F (2015) Ecosystem simplification, biodiversity loss and plant virus emergence. Curr. Opin. Virol. 10, 56-62. https://doi.org/10.1016/j.coviro.2015.01.005

Roossinck MJ, Martin DP, Roumagnac P (2015) Plant virus metagenomics: advances in virus discovery. Phytopathology, 105, 716-727. https://doi.org/10.1094/PHYTO-12-14-0356-RVW

Roossinck MJ, Saha P, Wiley GB, Quan J, White JD, Lai H, Chavarrıa F, Shen G, Roe BA (2010) Ecogenomics: using massively parallel pyrosequencing to understand virus ecology. Mol. Ecol. 19, 81-88. https://doi.org/10.1111/j.1365-294X.2009.04470.x

Rozado-Aguirre Z, Adams I, Collins L, Fox A, Dickinson M, Boonham N (2016) Detection and transmission of Carrot torrado virus, a novel putative member of the Torradovirus genus. J. Virol. Meth. 235, 119-124. https://doi.org/10.1016/j.jviromet.2016.05.018

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol. Biol. Evol. 34, 3299-3302. https://doi.org/10.1093/molbev/msx248

Sallinen S, Norberg A, Susi H, Laine AL (2020) Intraspecific host variation plays a key role in virus community assembly. Nat. Comm. 11,5610. https://doi.org/10.1038/s41467-020-19273-z

Schönegger D, Babalola BM, Marais A, Faure C, Candresse T (2022) Diversity of polerovirus-associated RNAs in the virome of wild and cultivated carrots. Plant Pathol 71, 1892– 1900. https://doi.org/10.1111/ppa.13623

Stubbs L. L. (1948) .-A new virus disease of carrots: its transmission, host range, and control. Aust. J. Sci. Res. B 1: 303-32.

Susi H, Filloux D, Frilander MJ, Roumagnac P, Laine AL (2019) Diverse and variable virus communities in wild plant populations revealed by metagenomic tools. Peer J. 7, e6140. https://doi.org/10.7717/peerj.6140

Susi H, Laine AL (2021) Agricultural land use disrupts biodiversity mediation of virus infections in wild plant populations. New Phytol. 230, 2447-2458. https://doi.org/10.1111/nph.17156

Syller J (2003) Molecular and biological features of umbraviruses, the unusual plant viruses lacking genetic information for a capsid protein. Physiol. Molec. Plant Pathol. 63, 35-46. https://doi.org/10.1016/j.pmpp.2003.08.004

Thapa V, McGlinn DJ, Melcher U, Palmer MW, Roossinck MJ (2015) Determinants of taxonomic composition of plant viruses at the Nature Conservancy's Tallgrass Prairie Preserve, Oklahoma. Virus Evol. 1, vev007. https://doi.org/10.1093/ve/vev007

Tugume AK, Mukasa SB, Valkonen JPT (2016) Mixed infections of four viruses, the incidence and phylogenetic relationships of sweet potato chlorotic fleck virus (*Betaflexiviridae*) isolates in wild species and sweetpotatoes in Uganda and evidence of distinct isolates in East Africa. PLOS One 11, e0167769. https://doi.org/10.1371/journal.pone.0167769

van der Valk T, Vezzi F, Ormestad M, Dalén L, Guschanski K (2019) Index hopping on the Illumina HiseqX platform and its consequences for ancient DNA studies. Mol. Ecol. Resour. 20, 1171-1181 https://doi.org/10.1111/1755-0998.13009

Veech JA (2013) A probabilistic model for analysing species co-occurrence. Glob. Ecol. Biogeogr. 22, 252-260. https://doi.org/10.1111/j.1466-8238.2012.00789.x

Williams IS, Dixon AFG (2007) Life cycles and polymorphism. In Aphid as crop pests, van Emden HF, Harrington R (eds), CABI, Wallington (UK), pp 69-85.

Wu H, Fu W, Fu Q, Yu F, Peng Y (2022) Plant Virus Database: a resource for exploring the diversity of plant viruses and their interactions with hosts. *bioRxiv*. 485054, https://doi.org/10.1101/2022.03.20.485054

Yoshida N (2020) Biological and genetic characterization of carrot red leaf virus and its associated virus/RNA isolated from carrots in Hokkaido, Japan. Plant Pathol. 69, 1379-1389. https://doi.org/10.1111/ppa.13202

Table 1. Known and novel viruses and virus-like agents identified in carrot populations sampled in 2019-2020 in Nouvelle-Aquitaine (France).

Virus family	Virus genus	Virus name	Virus acronym	Known or novel viruses	Category	Accession no.
Solemoviridae	*Polerovirus*	carrot red leaf virus	CtRLV	Known	(+)ssRNA	NC_006265
Solemoviridae	*Polerovirus*	wild carrot red leaf virus	WCtRLV			LT615231
Tombusviridae	*Umbravirus*	carrot mottle virus	CMoV			KF533714
Tombusviridae	*Umbravirus*	wild carrot mottle virus	WCMoV			LT615232
Tombusviridae	*Umbravirus*	carrot mottle mimic viurs	CMoMV			NC_001726
Closteroviridae	*Closterovirus*	carrot yellow leaf virus	CYLV			KF533699
Betaflexiviridae	*Chordovirus*	carrot Ch virus 1	CChoV1			NC_025469
Betaflexiviridae	*Chordovirus*	carrot Ch virus 2	CChoV2			NC_025468
Secoviridae	*Torradovirus*	carrot torradovirus 1	CaTV1			NC_025479 and NC_025480
Bromoviridae	*Ilarvirus*	solanum nigrum ilarvirus 1	SnIV1			MN216370
Partitiviridae	*Alphapartitivirus*	carrot cryptic virus	CaCV		dsRNA	NC_038824 and NC_038823
Partitiviridae	*Betapartitivirus*	dill cryptic virus 2	DiCV2			NC_021147 and NC_021148
unclassified	*unclassified*	carrot mottle virus satellite RNA	CMoV-satRNA		linear ss satellite RNA	NC_030649
unclassified	*unclassified*	carrot red leaf virus associated RNA	CtRLVaRNA		Tombus-like associated RNA	NC_003871
unclassified	*unclassified*	arracacha latent virus E associated RNA	ALEVaRNA			MF136436
unclassified	*unclassified*	beet western yellows virus associated RNA	BWYVaRNA			KF533709
unclassified	*unclassified*	carrot umbra-like virus 1	CaULV1			OP889249
Benyviridae	*unclassified*	carrot associated RNA virus 1	CaRNAV1		unknown	OP889248

Solemoviridae	*Polerovirus*	carrot polerovirus 1	CaPV1	Novel	(+)ssRNA	OP886450
Solemoviridae	*Polerovirus*	carrot polerovirus 2	CaPV2			OP886451
Solemoviridae	*Enamovirus*	carrot enamovirus 1	CaEV1			OP886449
Tombusviridae	*Umbravirus*	carrot umbravirus 1	CaUV1			OP886454
Tombusviridae	*Umbravirus*	carrot umbravirus 2	CaUV2			OP886452
Tombusviridae	*Umbravirus*	carrot umbravirus 3	CaUV3			OP886453
Closteroviridae	*Closterovirus*	carrot closterovirus 2	CtCV2			OP886455
Closteroviridae	*Closterovirus*	carrot closterovirus 3	CtCV3			OP886456
Betaflexiviridae	*Chordovirus*	carrot chordovirus 3	CChV3			OP886457
Partitiviridae	*Deltapartitivirus*	carrot cryptic virus 2	CaCV2		dsRNA	OP886461 and OP886462
Totiviridae	*Totivirus*	carrot associated toti-like virus	CaaTLV			OP886480
Amalgaviridae	unclassified	carrot amalgavirus 1	CaAV1			OP886466
unclassified	unclassified	carrot red leaf virus associated RNA 2	CtRLVaRNA2		Tombus-like associated RNA	OP889248
unclassified	unclassified	carrot mottle virus satellite RNA 2	CMoVsatRNA2		linear ss satellite RNA	OP886481
unclassified	unclassified	carrot red leaf virus satellite 1	CtRLV Sat1		small circular ss satellite RNAs	OM962993
unclassified	unclassified	carrot red leaf virus satellite 2	CtRLV Sat2		small circular ss satellite RNAs	OM962994
unclassified	unclassified	carrot flavi-like virus 1	CtFLV1		unknown	OM681407.1
unclassified	unclassified	carrot Ker-like virus	CaKLV			to be published elsewhere (TC, AM, DS et al., manuscript in preparation)

.

Table 2.: Prevalence of known and novel aphid-borne viruses and virus-like agents in different carrot populations. The number of plants individually tested by RT-PCR for each population is indicated, together with the populations code (See Supplementary Table 1). Viral acronyms are according to Table 1. Cells are color coded with the intensity of the red shading corellated with viral prevalence.

		Population type and code number							
		Cultivated		**Off-type**	**Wild**				
		19_1 *(49 plants)*	**19_2** *(50 plants)*	**19_7** *(39 plants)*	**19_6** *(50 plants)*	**19_9** *(44 plants)*	**19_12** *(50 plants)*	**19_13** *(50 plants)*	**19_14** *(44 plants)*
Known viruses and satellites	**CtRLV**	**100%**	**100%**	**100%**	**24%**	**25%**	**20%**	**42%**	**14%**
	CMoV	**80%**	**100%**	**100%**	**16%**	**20%**	**18%**	**64%**	**20%**
	CMoMV	**10%**	**70%**	**56%**	**10%**	**9%**	**16%**	**30%**	**23%**
	WCMoV	**20%**	**88%**	**77%**	**14%**	**36%**	**20%**	**34%**	**23%**
	CtRLVaRNA	**86%**	**98%**	**100%**	**16%**	**9%**	**22%**	**54%**	**27%**
	ALVEaRNA	**29%**	**98%**	**95%**	**2%**	0%	0%	**10%**	0%
	BWYVaRNA	0%	0%	**3%**	0%	0%	0%	0%	0%
	CMoV-satRNA	**33%**	**96%**	**95%**	**18%**	**23%**	**40%**	**32%**	**16%**
Novel viruses and satellites	**CaUV1**	**12%**	**12%**	**3%**	0%	0%	0%	0%	0%
	CaUV2	**2%**	**2%**	0%	**4%**	**7%**	**14%**	**18%**	**14%**
	CaUV3	0%	0%	0%	**2%**	**5%**	**8%**	0%	0%
	CtRLVaRNA2	0%	0%	**3%**	**2%**	**11%**	0%	2%	**14%**
	CMoV satRNA-2	0%	**2%**	0%	**2%**	**5%**	**10%**	**20%**	**5%**
	CaPV1	0%	0%	0%	**6%**	**20%**	**6%**	0%	**16%**
	CaPV2	**2%**	0%	0%	0%	0%	0%	0%	0%
	CaEV1	**4%**	0%	0%	**6%**	**9%**	**4%**	**14%**	**14%**

Table 3: Genetic diversity of different aphid-transmitted viruses and virus-like agents in different types of carrot populations. The values shown, together with their standard deviation, were calculated from multiple alignments of RT-PCR amplicons and HTS contigs sequences from individually analysed plants. P-values were calculated using a Mann-Whitney-Wilcoxon non parametric test.

	Carrot population type			**p-value***		
	Cultivated	**Off-type**	**Wild**	**Cultivated/wild**	**Cultivated/off-type**	**Off-type/wild**
CtRLV	1.00% ± 0.31%	1.73% ± 0.46%	5.71% ± 0.87%	**0.000**	**0.001**	**0.000**
CMoV	1.43% ± 0.52%	1.40% ± 0.52%	1.80% ± 0.58%	**0.000**	**0.951**	**0.014**
CMoMV	3.73% ± 0.46%	6.73% ± 0.82%	8.19% ± 0.89%	**0.000**	**0.002**	**0.131**
WCMoV	6.88% ± 0.79%	4.08% ± 0.22%	3.15% ± 0.46%	0.648	0.986	**0.002**
CMoVsatRNA	2.33% ± 0.63%	2.42% ± 0.67%	2.29% ± 0.67%	0.209	0.951	0.575
CtRLVaRNA	7.17% ± 0.87%	6.66% ± 0.93%	8.09% ± 0.91%	0.52	0.054	0.504

* p-value calculated to evaluate significance of a difference in diversity between isolates from cultivated and from wild carrots.

Figure 1. Map showing the carrot sampling sites in the two departments of Gironde and Landes of the Nouvelle-Aquitaine region of France. In the close up at the left, red balloons correspond to sampling site of cultivated and/or off-type carrots while green balloons correspond to sampling sites of wild populations

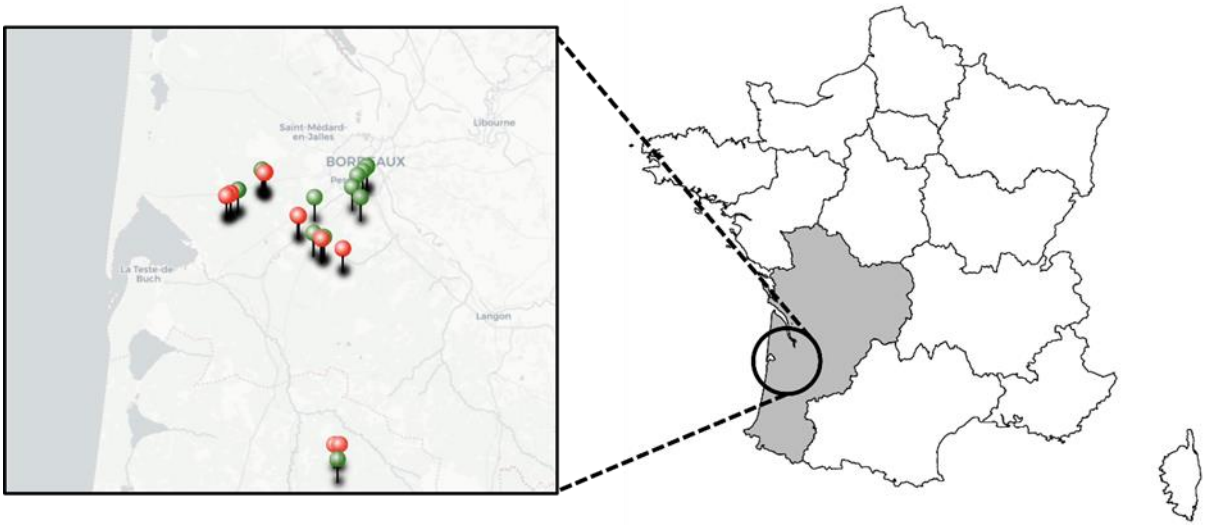


Figure 2: Heatmap showing a hierarchical clustering based on the number of mapped reads against viral references for carrot populations sampled in 2019 and 2020. Viral reference sequences are listed as acronyms (according to Table 1) in the rows on the right of the heatmap. The different carrot populations are represented in the columns and their coding at the bottom is according to Table S1 with suffixes indicating the type of population: c = cultivated field; cb = border of a field cc: center of a field; off = off-type carrots within cultivated carrot fields; w = wild population. The field compartment (cultivated + off-type) is represented by the red branches while the wild compartment is indicated by green branches. The color of the matrix cells reflects the number of reads, following the Z-score scale at the upper left.

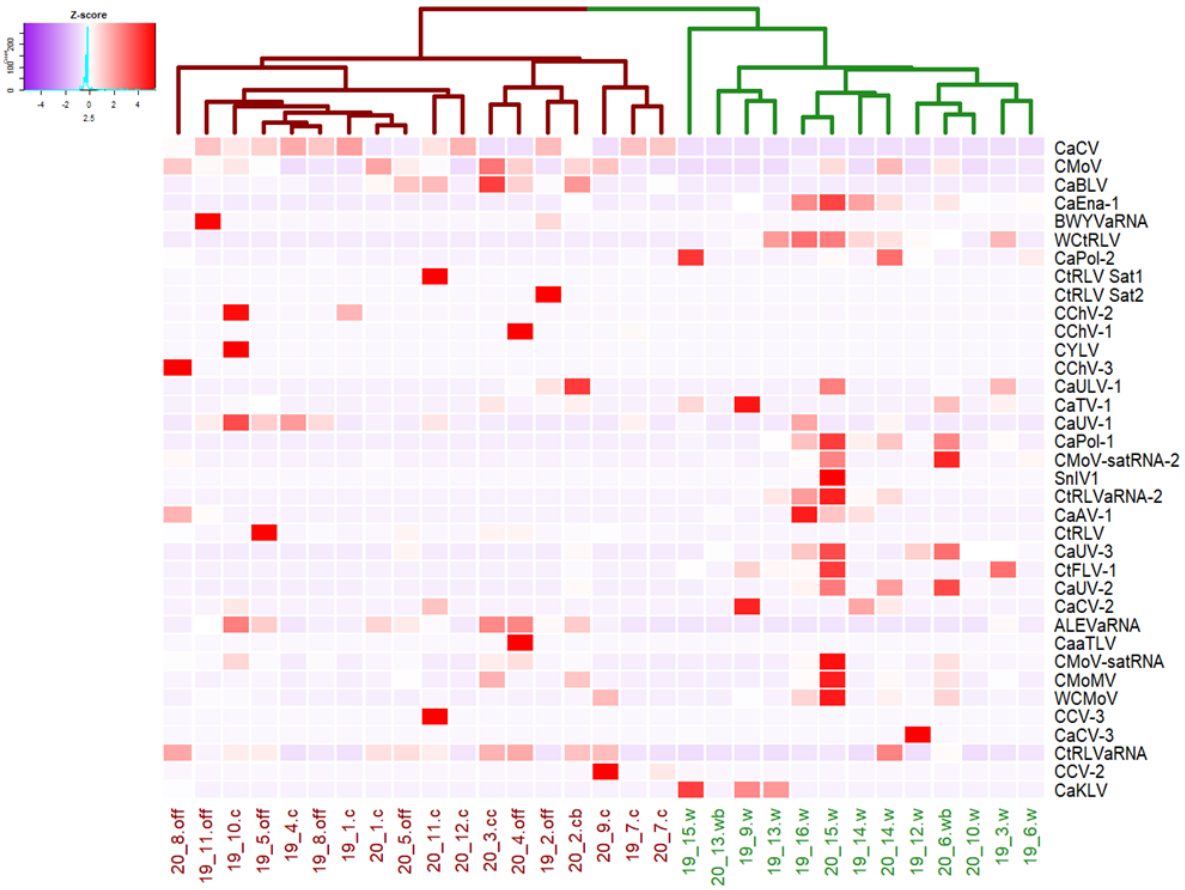


Figure 3: Venn diagram showing the number and corresponding percentages of viruses identified in the carrot virome that are exclusively found in either cultivated, off-type or wild populations and the number of viruses shared amongst the different population types

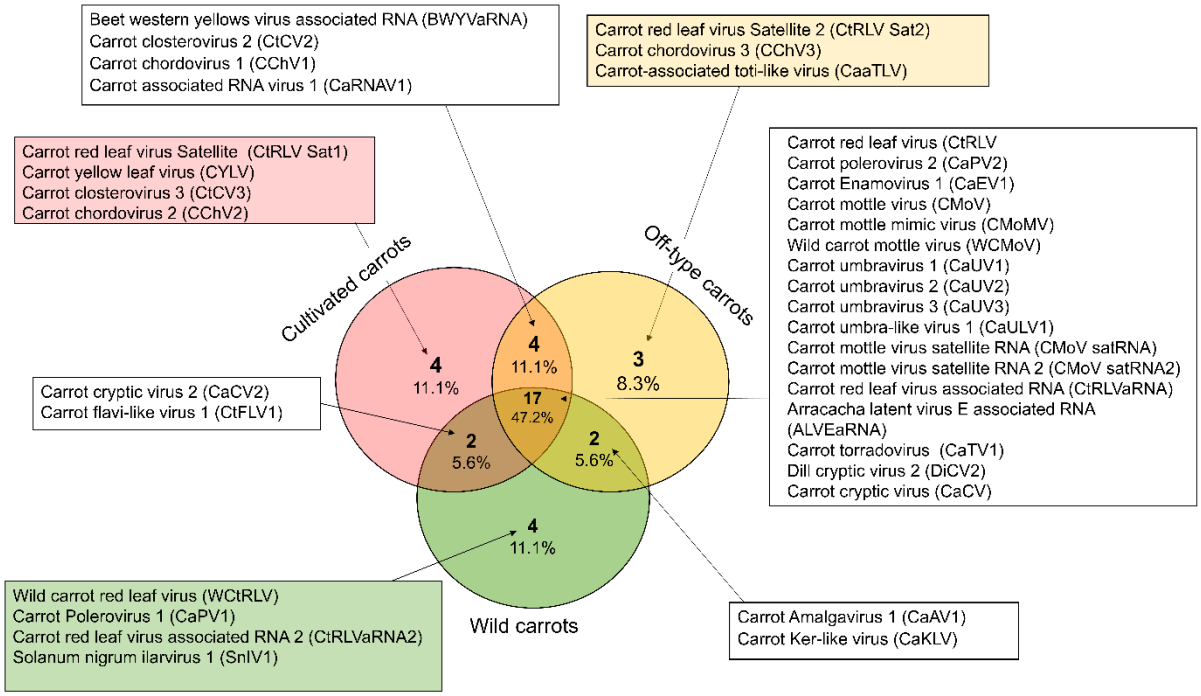
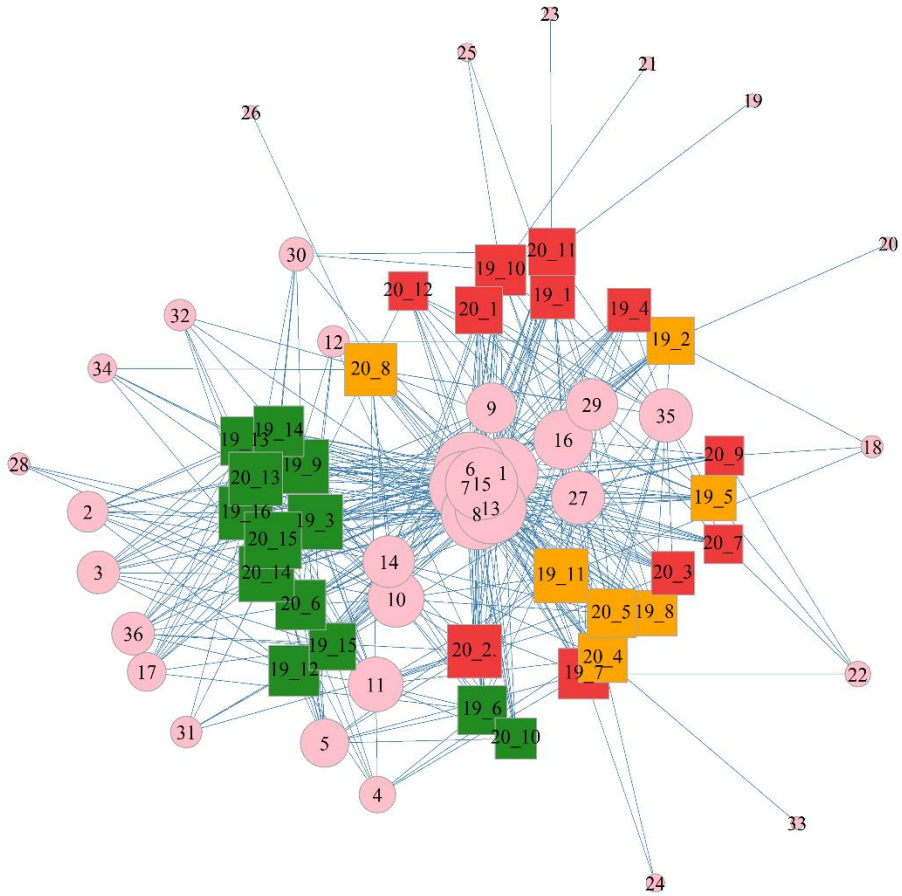


Figure 4: Bipartite network based on a binary incidence matrix derived from the reads mapping matrix. The network consists of two types of nodes: circles representing viruses and squares representing the carrot populations colored according to their type: green, wild; orange, off-type; and red, cultivated populations. The coding of the host populations is based on Table S1, the coding of viral nodes is given on the right of the network graph with viral acronyms according to Table 1. The size of the viral nodes corresponds to their degree (number of host populations in which viruses were detected).



node number	virus species
1	CtRLV
2	WCtRLV
3	CaPV1
4	CaPV2
5	CaEV1
6	CMoV
7	WCMoV
8	CMoMV
9	CaUV1
10	CaUV2
11	CaUV3
12	CaULV1
13	CMoV-satRNA
14	CMoVsatRNA2
15	CtRLVaRNA
16	ALEVaRNA
17	CtRLVaRNA2
18	BWYVaRNA
19	CtRLV Sat1
20	CtRLV Sat2
21	CYLV
22	CtCV2
23	CtCV3
24	CChV-1
25	CChV-2
26	CChV-3
27	CaTV1
28	SnIV1
29	CaCV
30	DiCV2
31	CaCV2
32	CaAV-1
33	CaaTLV
34	CaKLV
35	CaRNAV1
36	CtFLV1

Figure 5: Heatmap showing co-occurrence associations between pairs of viruses or viral agents. Associations are colour coded according as follow: Positive, light blue; random, grey; negative, yellow. Helper viruses of the *Solemoviridae* family are indicated in pink, dependent umbraviruses in blue, and dependent subviral agents in green. P-values indicating statistical significant non-random associations are given in the cells of the heatmap. B) Pie-chart showing the percentage of identified types of positive associations broken down by virus types: D-D: associations between dependent umbraviruses; D-d, associations between umbraviruses and subviral agents, H-D, associations between a *Solemoviridae* helper virus and a dependent umbravirus; H-d association between a *Solemoviridae* helper virus and a subviral agents, H-H associations between different *Solemoviridae* helper viruses.

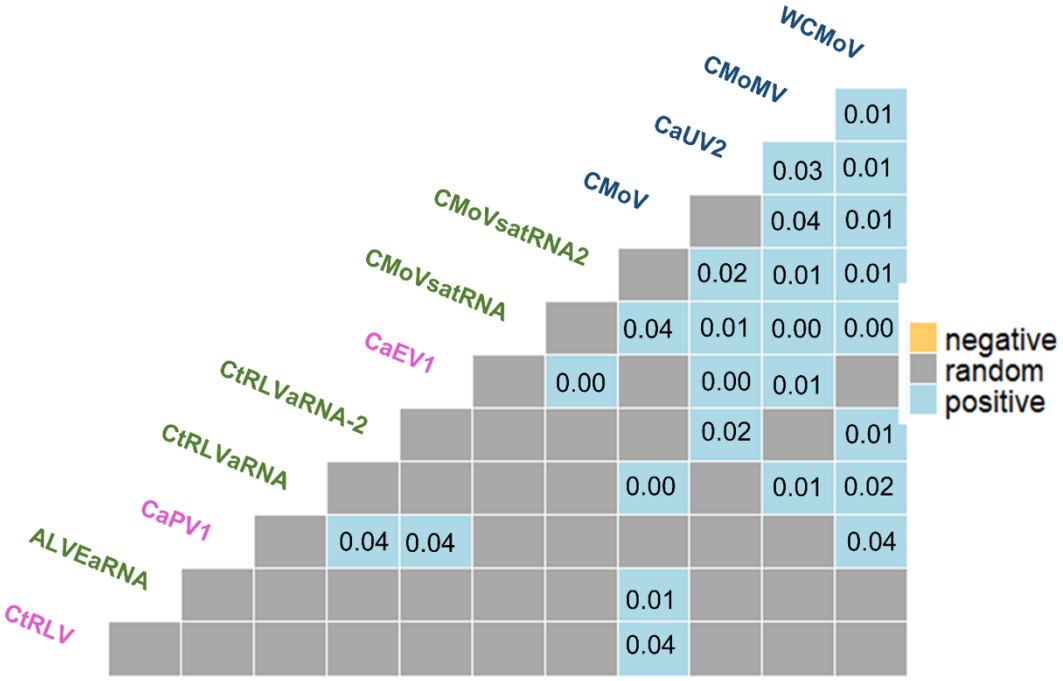
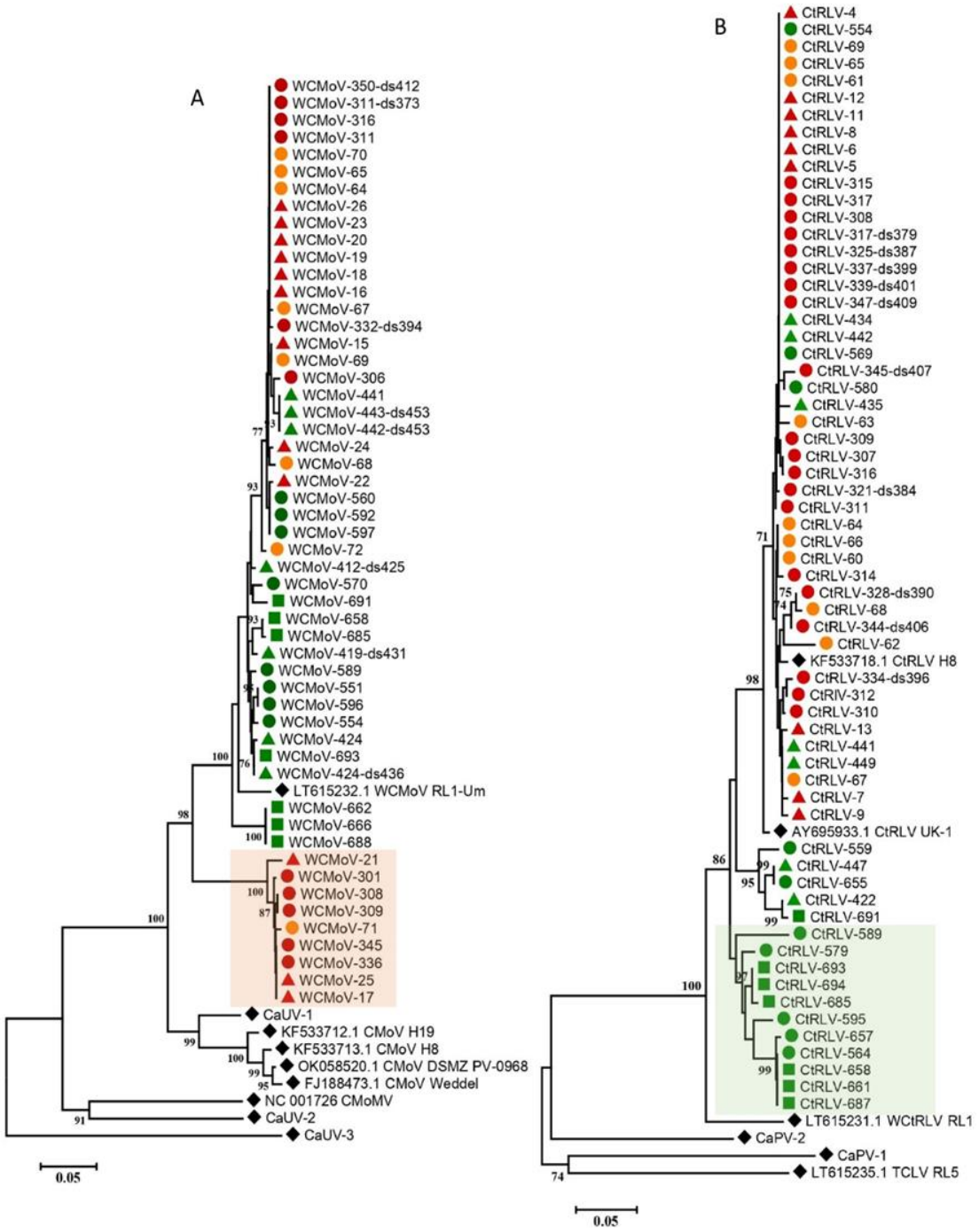


Figure 6: Neighbour joining phylogenetic trees reconstructed using multiple nucleotide sequence alignments of RT-PCR amplicons and HTS contigs from individually analysed plant of (A) WCMoV and (B) CtRLV. Colors indicate the type of carrot population: red, cultivated; orange, off-type; and green, wild. Only bootstrap values higher than 70% are indicated. Different sampling sites are indicated by different shapes. Reference sequences are indicated by black diamonds. Specific subclusters from cultivated or wild populations are indicated by red and green shaded boxes, respectively.



Supplementary Materials [Supplementary Table S4 that does not fit easily into an A4 format is available at https://doi.org/10.57745/NAV8II]

Supplementary Table S1 : information on sampled carrot populations in Nouvelle-Aquitaine (France) over 2019-2020

Supplementary Table S2 : RT-PCR primer pairs targenting aphid-transmitted viruses of the families Tombusviridae and Solemoviridae and associated subviral agents

Supplementary Table S3 : Number of infected carrot populations per population type (cultivated, off-type, wild) for all viruses identified in the Nouvelle Aquitaine carrot virome. Viruses are grouped according to their identification in different carrot population types.

Supplementary Table S4 : Results for the co-occurrence analysis presented in Figure 5 of the main text. Positive associations between virus pairs are highlighted in red

Supplementary FigureS1: Neighbour joining phylogenetic trees reconstructed using multiple nucleotide sequence alignments of RT PCR amplicons and HTS contigs of (A) CtRLVaRNA and (B) CMoV. Colours indicate the type of carrot population from which the sequence was obtained: red: cultivated; orange: off-type and green: wild. Different sampling sites are indicated by different shapes. Reference sequences are indicated by black diamonds

Supplementary Figure S2: Neighbour joining trees reconstructed from RT-PCR amplicon sequences of CtRLVaRNA (A), CMoVsatRNA2 (B), CaEV1 (C), CaPV1 (D), CaUV1 (E), CaUV2 (F), CaUV3 (G). Isolates of wild carrots, off-type and cultivated carrots are indicated in green, orange and red, respectively. Different shapes are used to indicate different populations/sampling sites. Reference sequences are indicated with black diamonds.

Supplementary Table S1

Population-ID	Population type	Region/province	GPS point	Date of sampling
19_1	cultivated carrot	Nouvelle-Aquitaine/Landes	44°16'10.0"N 0°39'44.6"W	30/07/2019
19_2	wild off-type carrot	Nouvelle-Aquitaine/Landes	44°16'10.0"N 0°39'44.6"W	30/07/2019
19_3	wild carrot	Nouvelle-Aquitaine/Landes	44°14'25.696"N 0°39'14.5"W	30/07/2019
19_4	cultivated carrot	Nouvelle-Aquitaine/Gironde	44°44'07.7"N 0°56'41.2"W	30/07/2019
19_5	wild off-type carrot	Nouvelle-Aquitaine/Gironde	44°44'07.7"N 0°56'41.2"W	30/07/2019
19_6	wild carrot	Nouvelle-Aquitaine/Gironde	44°44'43.539"N 0°54'53.284"W	30/07/2019
19_7	cultivated carrot	Nouvelle-Aquitaine/Gironde	44°41'51.1"N 0°45'29.5"W	30/07/2019
19_8	wild off-type carrot	Nouvelle-Aquitaine/Gironde	44°41'51.1"N 0°45'29.5"W	30/07/2019
19_9	wild carrot	Nouvelle-Aquitaine/Gironde	44°43'55.07"N 0°42'51.001"W	30/07/2019
19_10	cultivated carrot	Nouvelle-Aquitaine/Gironde	44°38'10.6"N 0°38'36.4"W	30/07/2019
19_11	wild off-type carrot	Nouvelle-Aquitaine/Gironde	44°38'10.6"N 0°38'36.4"W	30/07/2019
19_12	wild carrot	Nouvelle-Aquitaine/Gironde	44°39'27.658"N 0°42'40.752"W	30/07/2019
19_13	wild carrot	Nouvelle-Aquitaine/Gironde	44°44'49.484"N 0°37'2.188"W	22/08/2019
19_14	wild carrot	Nouvelle-Aquitaine/Gironde	44°43'49.625"N 0°35'31.52"W	22/08/2019
19_15	wild carrot	Nouvelle-Aquitaine/Gironde	44°46'4.296"N 0°36'26.916"W	22/08/2019
19_16	wild carrot	Nouvelle-Aquitaine/Gironde	44°47'20.76"N 0°34'32.753"W	22/08/2019
20_1	cultivated carrot	Nouvelle-Aquitaine/Gironde	44°46'51.4"N 0°51'02.9"W	01/07/2020
20_2	cultivated carrot-field edge	Nouvelle-Aquitaine/Gironde	44°46'51.4"N 0°51'02.9"W	01/07/2020
20_3	cultivated carrot-field centre	Nouvelle-Aquitaine/Gironde	44°46'51.4"N 0°51'02.9"W	01/07/2020
20_4	wild off-type carrot	Nouvelle-Aquitaine/Gironde	44°46'51.4"N 0°51'02.9"W	01/07/2020
20_5	wild off-type carrot	Nouvelle-Aquitaine/Gironde	44°46'51.4"N 0°51'02.9"W	01/07/2020
20_6	wild carrot-bordering carrot field	Nouvelle-Aquitaine/Gironde	44°46'51.4"N 0°51'02.9"W	01/07/2020
20_7	cultivated carrot	Nouvelle-Aquitaine/Gironde	44°44'20.8"N 0°55'59.3"W	08/07/2020
20_8	wild off-type carrot	Nouvelle-Aquitaine/Gironde	44°44'20.8"N 0°55'59.3"W	08/07/2020
20_9	cultivated carrot	Nouvelle-Aquitaine/Gironde	44°44'20.8"N 0°55'59.3"W	08/07/2020
20_10	wild carrot	Nouvelle-Aquitaine/Gironde	44°44'43.5"N 0°54'55.0"W	08/07/2020
20_11	cultivated carrot	Nouvelle-Aquitaine/Gironde	44°39'10.5"N 0°42'22.1"W	08/07/2020
20_12	cultivated carrot	Nouvelle-Aquitaine/Gironde	44°39'10.5"N 0°42'22.1"W	08/07/2020
20_13	wild carrot- bordering carrot field	Nouvelle-Aquitaine/Gironde	44°39'10.5"N 0°42'22.1"W	08/07/2020
20_14	wild carrot	Nouvelle-Aquitaine/Gironde	44°46'46.4"N 0°35'15.0"W	08/07/2020
20_15	wild carrot	Nouvelle-Aquitaine/Gironde	44°47'16.3"N 0°34'34.9"W	08/07/2020
20_16	wild carrot	Nouvelle-Aquitaine/Gironde	48°36'47.2"N 1°54'18.3"E	13/07/2020

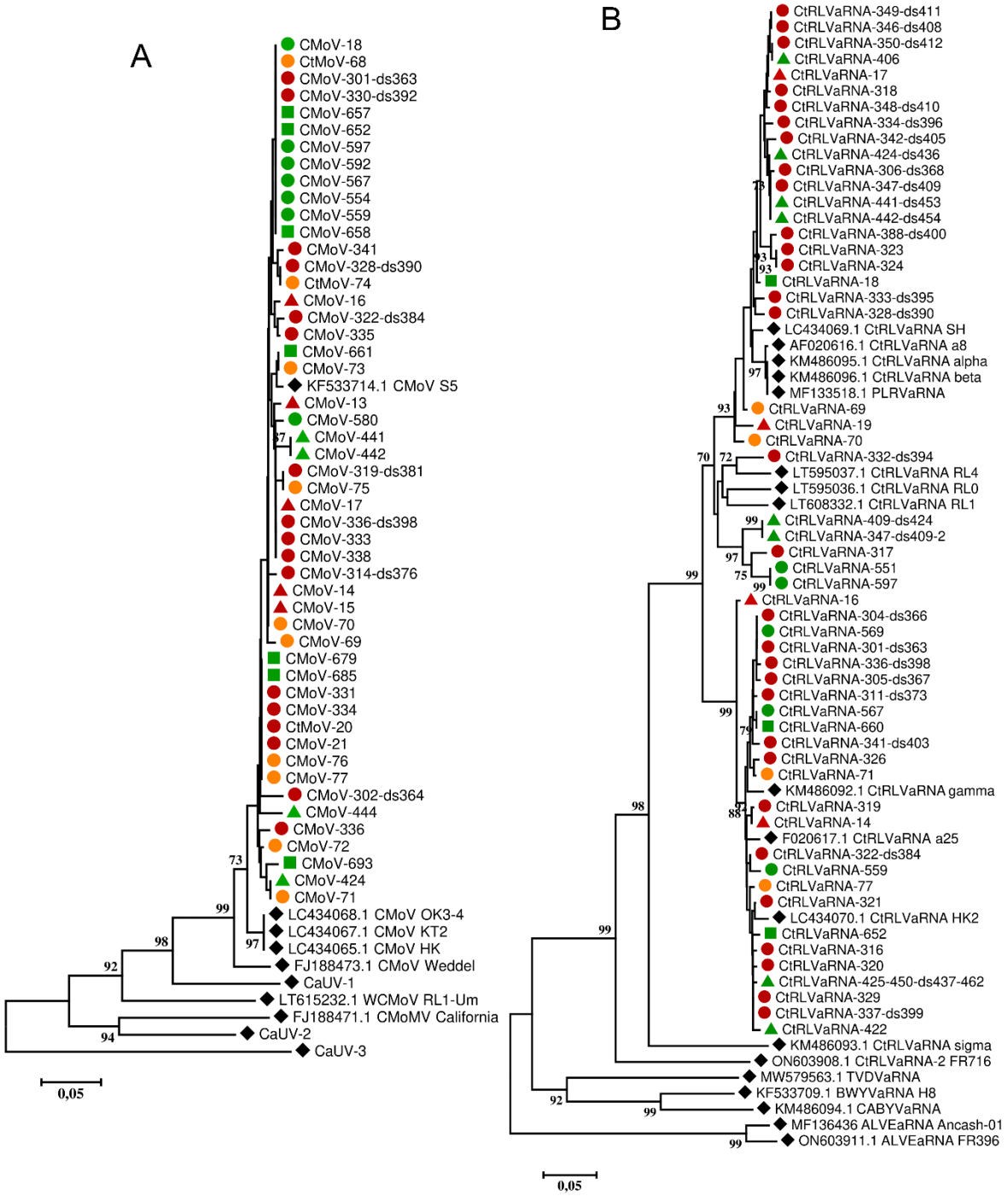
Supplementary Table S2

Family	Genus	virus name	Acronym	Primer name	5' - 3' sequence	Amplicon length	Annealing temperature in °C	amplified gene
Tombusviridae	Umbravirus	carrot mottle mimic virus	CMoMV	CtMoMiV-F1 CtMoMiV-R1	CTCCTTGGTGAATCTCGTGAG GGCATTAAATCCCTTGGCGAC	438 bp	58	RdRP
	Umbravirus	carrot mottle virus	CMoV	CtMoV-F1 CtMoV-R1	GAATTGAGCAAGACAGCATCT AYCACTGCYTGTCCCGGCGA	289 bp	55	ORF3 (MP-protein)
	Umbravirus	wild carrot mottle virus	WCMoV	WCtMoV-F1 WCtMoV-R1	GAGCTAGGCCTCGGARTYATGAC TCCTCCGGTAGTGTYGGCYTG	499 bp	60	cell to cell movement protein
	Umbravirus	carrot umbravirus 1	CaUV1	Umbra_1-F1 Umbra_1-R1	GATGACCAGAGCACTATGCCTTGA GGATCCCAGGTGGTCGTATGCGTT	576 bp	65	RdRP
	Umbravirus	carrot umbravirus 2	CaUV2	Umbra_2-F1 Umbra_2-R1	GTGTTGTATTTATGGAACGCC GATCTTCAATGTGCATTCCCTT	407 bp	55	RdRP
	Umbravirus	carrot umbravirus 3	CaUV-3	Umbra_3-F1 Umbra_3-R1	TGTCAACAGTGATAAGAAAAGCTC CCTCTGTAGGTAGTTGTGGCTCA	569 bp	62	cell to cell movement protein
Solemoviridae	Polerovirus	carrot red leaf virus	CtRLV	CtRLV-F2 CtRLV-R2	GGGCTTTTAACAGCTTAYCAC TCTAACACGGACAAGGGRGAT	263 bp	55	P1-Protein
	Polerovirus	carrot polerovirus 1	CaPV1	CaPol1-F1 CaPol1-R1	GCGTGTGAGGGATTTGTAACCT CCTCTAGGATTTTGCCCATCAT	885 bp	58	P1-Protein
	Polerovirus	carrot polerovirus 2	CaPV2	CaPol2-F1 CaPol2-R1	CCTGTGAAGGTTTTATCACATA GGTCCCATATAAAGGACCAGGTC	267 bp	58	P1-Protein
	Enamovirus	carrot enamovirus 1	CaEV1	CaEna1-F1 CaEna1-R1	TTGTGACTGTTGAGGCACTAATG CACATCCTTTGCTCATTGGCAAA	402 bp	60	P1-Protein
Satellites	unclassified	carrot mottle virus satellite RNA	CtMoV-satRNA	CtMoV-sat-F1 CtMoV-sat-R1	TAGACGACATGTCTGCTGTCCC CATCCGTAGTCAAATCCTGCT	324 bp	60	
	unclassified	carrot red leaf virus associated RNA	CtRLVaRNA	CtRLVaRNA-F1 CtRLVaRNA-R1	GAATGGACCAAGAGCTTCTCG ATTGGCTACTTCAATGTGTATG	425 bp	62	RdRp
	unclassified	carrot mottle virus satellite RNA 2	CtMoVsatRNA2	div-sat-F1 div-sat-R1	CGTGGGAAAAGATGCTCTGGTT ATATTGGTCTCGCTGCCTTAG	236 bp	60	
	unclassified	arracacha latent virus E associated RNA	ALVEaRNA	ALVEaRNA-F1 ALVEaRNA-R1	CTTTCTTTAGGAGTAGTGAGACC GAACTGTAGGGCATCACAACCT	606 bp	60	RdRp
	unclassified	carrot red leaf virus associated RNA 2	Divergent aRNA	div-aRNA-F1 div-aRNA-R1	CGCTTGGGATAGTTTGAGAGAC GATCGTAGCCATGAAACGCTTG	547 bp	60	RdRp
	unclassified	beet western yellows virus associated RNA	BWYVaRNA	BWYVaRNA-F1 BWYVaRNA-R1	ATGCACTCCAGCATACCCACTT CATCGTGCATACATCAGGAGTGA	444 bp	58	RdRp

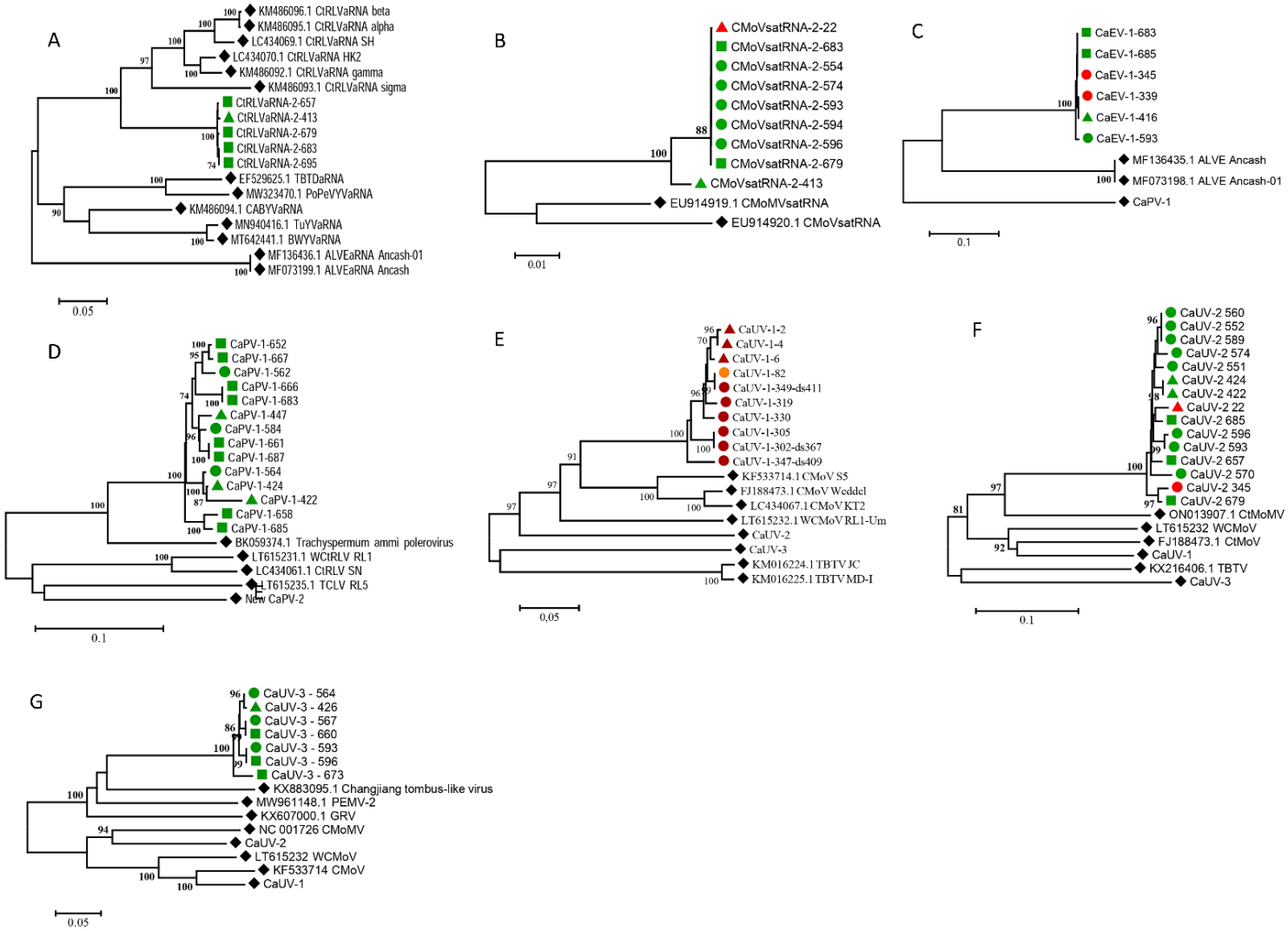
Supplementary Table 3:

		Number of infected populations per population type					
		2019			2020		
differential or shared presence	virus	cultivated	off-type	wild	cultivated	off-type	wild
exclusively present in cultivated populations	CtRLV Sat1	0/4	0/4	0/8	**1/7**	0/3	0/5
	CYLV	**1/4**	0/4	0/8	0/7	0/3	0/5
	CtCV3	0/4	0/4	0/8	**1/7**	0/3	0/5
	CChV2	**2/4**	0/4	0/8	0/7	0/3	0/5
exclusively present in off type populations	CtRLV Sat2	0/4	**1/4**	0/8	0/7	0/3	0/5
	CtChV3	0/4	0/4	0/8	0/7	**1/3**	0/5
	CaaTLV	0/4	0/4	0/8	0/7	**1/3**	0/5
exclusively present in wild populations	WCtRLV	0/4	0/4	**6/8**	0/7	0/3	**5/5**
	CaPol-1	0/5	0/4	**7/8**	0/7	0/3	**4/5**
	CtRLVaRNA2	0/4	0/4	**6/8**	0/7	0/3	**2/5**
	SnIV1	0/4	0/4	**1/8**	0/7	0/3	**2/5**
viruses common in cultivated and off-type	BWYVaRNA	0/4	**2/4**	0/8	**1/7**	0/3	0/5
	CtCV2	**1/4**	**1/4**	0/8	**2/7**	0/3	0/5
	CChV1	**1/4**	0/4	0/8	0/7	**1/3**	0/5
	CaRNAV1	**4/4**	**4/4**	0/8	**7/7**	**3/3**	0/5
viruses common in wild and off-type	CaAV1	0/4	**1/4**	**2/8**	0/7	**1/3**	**2/5**
	CaKLV	0/4	0/4	**3/8**	0/7	**1/3**	**1/5**
viruses common in cultivated and wild	CaCV3	0/4	0/4	**2/8**	**2/7**	0/3	**2/5**
	CtFLV1	0/4	0/4	**6/8**	**1/7**	0/3	**4/5**
viruses shared among all populaiton types	CtRLV	**4/4**	**4/4**	**8/8**	**7/7**	**3/3**	**4/5**
	CaPV2	**1/4**	**1/4**	**4/8**	0/7	**1/3**	**2/5**
	CaEV1	**1/4**	**1/4**	**6/8**	**1/7**	0/3	**5/5**
	CMoV	**4/4**	**4/4**	**8/8**	**7/7**	**3/3**	**5/5**
	WCMoV	**4/4**	**4/4**	**8/8**	**7/7**	**3/3**	**5/5**
	CMoMV	**4/4**	**4/4**	**8/8**	**6/7**	**3/3**	**5/5**
	CaUV1	**3/4**	**3/4**	**3/8**	**4/7**	**1/3**	**1/5**
	CaUV2	**2/4**	**1/3**	**7/8**	**2/7**	**2/3**	**5/5**
	CaUV3	0/4	0/4	**7/8**	**2/7**	**3/3**	**5/5**
	CaULV1	0/4	**1/3**	**1/8**	**1/7**	**1/3**	**2/5**
	CMoVsatRNA	**4/4**	**4/4**	**8/8**	**6/7**	**3/3**	**5/5**
	CMoVsatRNA2	**1/4**	0/4	**6/8**	**2/7**	**3/3**	**4/5**
	CtRLVaRNA	**4/4**	**4/4**	**8/8**	**7/7**	**3/3**	**5/5**
	ALEVaRNA	**4/4**	**4/4**	**3/8**	**6/7**	**3/3**	**1/5**
	CaTV1	**3/4**	**4/4**	**4/4**	**3/7**	**2/3**	**1/5**
	CaCV2	**1/4**	**1/3**	**2/8**	**1/7**	**0/3**	**2/5**
	CaCV	**4/4**	**4/4**	0/8	**5/7**	0/3	**1/5**

Supplementary Figure S1



Supplementary Figure S2



GENERAL DISCUSSION

In the present work, two aspects of virome description were investigated, one methodological the other virological and ecological in nature. With respect to the methodological approach, we showed that the use of a large-scale synthetic phytovirus community is a powerful tool for benchmarking HTS strategies. Synthetic communities are indeed already being used in clinical or ecological settings, but such comparative studies are largely lacking in plant virology. We used synthetic phytoviral communities of varying complexities, including a complex, 60+ members community to analyze to what extent more or less complex viromes are correctly captured by the HTS approaches used. Thus, the performance of two different virus enrichment strategies, dsRNA and VANA could be compared at different community complexity levels. Three key results have been obtained: (i) community complexity correlates negatively with virome description performance, (ii) dsRNA approach provided in all comparisons a more complete description of the RNA virome but, as expected, performed poorly with DNA viruses and (iii) a very robust positive correlation between the sequencing depth and the completeness of virome description was identified.

The second virological/ecological task aimed to contrast the virome of cultivated and wild carrot populations and analyze viral fluxes between them. The following questions have been addressed: (i) how far can we extend the carrot virome beyond what is already known, (ii) how different or similar are the viromes of cultivated and wild carrots, (iii) how stable is the carrot virome over space and time? (iv) is there evidence for virus movement or for barriers to movement between wild and cultivated carrots? The carrot pathosystem was selected to address these questions because wild and cultivated carrots belong to the same species and thus present a system with very low botanical and genetic barriers to virus flow across the agroecological interface between a crop and its weedy relative. Based on the virome fingerprints identified in wild and field carrots, the results obtained raise two types of questions and perspectives.

1 Ecological and epidemiological questions and perspectives

We screened a total of 45 different wild and cultivated carrot populations from five different regions of France over a North-South transect and seven additional carrot populations from central Spain. A very rich and diverse virome was identified, with a total of 45 viral agents, more than half of them (25) being novel. We have thus largely expanded the existing global carrot virome, which previously comprised about 30 viruses according to public databases (Brunt et al., 1996; Brister et al., 2015; Mihara et al., 2016; Wu et al., 2022). Due to the difference in sampling effort, no absolute conclusions can be drawn regarding the differences

in viral richness between France and Spain. However, nearly all novel viruses (92%, 23/25) have been discovered in French carrot populations with only four novel viruses shared between both countries and another two novel viruses exclusively identified in Spanish carrots. Whether the particularly rich carrot virome identified in France is a unique feature, especially with regard to the Nouvelle-Aquitaine region (main French carrot production area from where most of our populations have been sampled) or whether a comparable or even a higher virus richness can be seen in other European and non-European countries remains an open question. **A parallel question is that of the overall, global richness of the carrot virome**: has a significant fraction of it already been identified or have we only touched the "tip of the iceberg" as the classical metagenomics metaphor would have it? In this regard, comparisons with the work carried out in the UK by InextVir partners and a collaboration with Prof. Emmanuel Geoffriau, coordinator of the French network of genetic resources “Carrot and other Daucus” (Briard et al., 2006) which allowed us to gain access to carrot populations from several additional European countries should allow us in the near future to further expand these carrot virome description efforts in Europe.

The virome of cultivated carrots was clearly distinct from that of wild carrots. This was seen both on a quantitative level using a clustering approach based on the number of mapped reads against reference sequences and on a qualitative level using bipartite networks based on binary, presence absence data. Both analyses confirmed a very contrasted virome between field and wild carrot populations. Two kinds of differences were identified: first, some viruses showed differential presence as they were either exclusively or predominantly present in wild or in the cultivated compartments and second, remarkable differences in prevalence level could be identified for some viruses between wild and cultivated carrots, as determined by HTS and RT-PCR analysis of individual plants. Concerning the first point, differentially present viruses were mostly (but not exclusively) persistent viruses such as carrot cryptic virus, which was identified in most field carrot populations but in only one wild population, or novel deltapartitiviruses and amalgaviruses found predominantly in wild carrots. These persistent viruses do not have any known vectors and are believed to be transmitted through seeds or pollen (Roossinck, 2012a), so that their differential presence can be understood as highlighting the limited genetic fluxes between cultivated and wild carrot populations. Concerning differences in prevalence of individual viruses, known aphid-transmitted viruses showed very high prevalence in field carrots while novel viruses, also expected to be aphid-transmitted showed a symmetrical

pattern, with intermediate levels in wild carrots but only very sporadic detection in cultivated carrots.

Through the analysis of the sequences obtained from individual plants (either through HTS or RT-PCR screening) virus populations could be further analysed, revealing for some of them additional intraspecific differentiation pattern suggestive of the absence or of incomplete virus flow between the wild and the cultivated compartements. This was seen for isolates of an umbravirus, wild carrot mottle virus (WCMoV), with a subpopulation of isolates specific of cultivated carrots and for isolates of a polerovirus, carrot red leaf virus (CtRLV), with a subpopulation of isolates specific of wild carrots. Such a host-specific differentiation is very surprising given that the remaining isolates of these two virus did not show the same pattern, suggesting that they are able to freely exchange between compartements. This raises the question of the **mechanism(s) limiting the movement of some isolates of aphid-transmitted viruses while other isolates or viruses seem to be able to move freely between compartments**. The underlying process(es) might be determined either by environmental dynamics such as the ecology and biology of aphids or by the genetic makeup of plant populations or by a combination of both. A potential approach to address this type of question would be to grow different genotypes of carrots (e.g. modern cultivated, ancestral cultivated and wild carrot genotypes) side by side, submitting them to the same environmental inoculation pressure before virome analysis to determine whether different host genotypes acquire similar or differentiated viral communities. This type of experiment is also termed “common garden” and can refer to different experimental designs in which plants with different known properties are grown under the same condition. In this way the genetic factors can be disentangled from environmental forces to analyse and quantify their contribution to the composition and structure of the virus communities that become associated with these plants in different settings. Such approaches have been extensively used in plant ecology (Thompson et al., 1991; Moloney et al., 2009; Berend et al., 2019; Malyshev et al. 2014), and can be used for a wide rangeof organisms, including animals (Rhymer, 1992), parasites and pathogens (Koskela, 2002; Laine, 2007; Sasu et al., 2009). An interesting example concerns the work and Sasu et al. (2009), who crossed a wild plant, *Cucurbita pepo* ssp. *texana*, with a related crop, *Cucurbita pepo* ssp. *ovifera* carrying a virus resistance transgene (VRT). The generated transgenic and non-transgenic populations were grown in the same fields and viral and bacterial diseases allowed to naturally develop. The authors tested the impact of the VRT on the fitness of wild plants, herbivory by cucumber beetles and incidence of viral and bacterial diseases and found that the

VRT provided a higher fitness to wild plants in terms of reproductive success but that this was at the cost of increased herbivory by the cucumber beetle who preferred healthy (mostly VRT) plants (Sasu et al., 2009). These results highlight the complex interplay between host genetics and environmental dynamics and their reciprocal effects on host fitness. In another work, clonal replicates of different *Plantago lanceolata* genotypes were placed in settings during seasonal virus epidemics (Sallinen et al., 2021). In this way, the role of host genotypes could be teased apart from environmental drivers (e.g. abiotic conditions, host population structure and history, vector communities etc.). Using joint species distribution modelling (JSDM) the authors identified that the host genotype was the most important determinant in virus community assembly in their system, explaining most of the variance in the data, followed by environmental factors in the local population context (Sallinen et al., 2021).

As part of the carrot pathosystem studied here, off-type carrots can be considered as a small-scale common garden experiment, since they were growing side by side with cultivated carrots but have a different genetic makeup. Our virome fingerprint analyses revealed that off-type carrots have a virome that shows many similarities with that of cultivated carrots in terms of virome composition but also in terms of prevalence of individual viruses. This suggests a major impact of the environment and local growth conditions on virus community assembly. However, interpreting off-type carrots as hybrids between cultivated and wild carrots means that they are genetically intermediate between wild and cultivated carrots. It is thus possible that some of the off-type carrots virome properties might result from their original, hybrid genetic makeup. This in turn limits our ability to draw a conclusion about which of the two factors, genetic or environmental, contributes the most in shaping the observed virus communities. Only carefully planned common garden experiments could allow us to cleanly unravel the individual contributions of these two important parameters.

2 Etiological and agro-economic questions and perspectives

An interesting aspect of the carrot virome highlighted in the present work is the large number of viruses and viral agents involved in complex interdependencies, some of which cause an important carrot disease with a complex etiology, carrot motley dwarf (CMD). Stubbs (1948) was the first to describe CMD in carrots in Australia and it became soon evident that the disease is caused by a mixed infection of two viruses, carrot red leaf virus (CtRLV, genus Polerovirus) and carrot mottle virus (CMoV, genus Umbravirus) (Watson et al., 1964). Infected crops exhibit foliage symptoms such as irregular chlorotic mottle and marginal reddening of lower leaves as

well as stunted growth and reduced seed production (Stubbs, 1948). This and subsequent studies showed that aphids transmit CtRLV from singly infected plants but that aphid transmission of CMoV requires coinfection with CtRLV and that the underlying process involves transcapsidation of CtMoV RNA by the CtRLV capsid protein (Elnagar and Mutant, 1978). Currently, carrot mottle virus (CMoV) is one of the 11 ICTV approved species within the *Umbravirus* genus. Umbraviruses represent an original genus of plant viruses that lack a capsid protein and therefore are unable to form transmissible virus particles (Syller, 2003). For this, they rely on coinfection with helper viruses from the *Solemoviridae* (formerly 0*Luteoviridae*) family which assist them for efficient aphid transmission through transcapsidation (Syller 2003). Many umbraviruses are involved in disease complexes with assistor viruses of the Polero- or Enamovirus genera, causing damage to important crops in different parts of the world (reviewed by Syller et al. 2003). An important example is groundnut rosette disease, causing considerable losses in peanut production in Africa (Naidu et al.1998). The disease is caused by groundnut rosette virus (GRV, genus *Umbravirus*), its assistor virus groundnut rosette assistor virus (GRAV, genus *Polerovirus*) and GRV satellite RNA (Hull and Adams, 1968; Naidu et al., 1998). The latter two are packaged into GRAV particles in order to be transmitted, mainly by the aphid *Aphis craccivora*. The satellite RNA plays an important role in the disease because it is responsible for symptom expression and required by GRV for aphid transmission (Naidu et al., 1998). Another important disease caused by a complex interplay of helper and dependent viruses is tobacco bushy top disease (TBTD) which causes significant damage to tobacco plants in different African and Asian countries (Chen et al., 2022). The disease complex consists of tobacco vein distorting virus (TVDV, genus *Polerovirus*), tobacco bushy top virus (TBTV, genus *Umbravirus*), tobacco bushy top virus satellite RNA (TBTVsatRNA) and TVDV-associated RNA (TVDVaRNA) (Mo et al., 2011; Chen et al., 2022).

Many biological, epidemiological and ecological studies on the CMD complex were performed in the second half of the 20th century (Stubbs, 1948, Watson and Serjeant, 1964; Murant, 1975, Waterhouse and Murant, 1983, Watson and Falk, 1994). More recently, two additional subviral agents have been described, carrot red leaf virus-associated RNA (CtRLVaRNA, Watson et al., 1998) and carrot mottle virus satellite RNA (CMoVsatRNA, Menzel et al., 2009), both associated with CMD. However, their etiological role in the disease remains unclear. A very recent study on the biology of CtRLV and CtRLVaRNA has started to shed light on the complex role the dependent CMoV and CtRLVaRNA and on their relationship with the helper virus CtRLV (Yoshida, 2021). The author found three important key results. Firstly, CMoV

moderately enhanced CtRLV accumulation and symptom intensity. Umbraviruses are indeed known to enhance and facilitate accumulation and systemic movement of coinfecting viruses (Ryabov et al. 1999 and 2001, Zhou et al., 2017). Secondly, CtRLVaRNA abolished the enhancement of CtRLV accumulation induced by CMoV and even decreased its accumulation to undetectable levels in some of the tested *Apiaceae* species. Thirdly, CtRLVaRNA (+CMoV) enhanced disease severity in carrots while it reduced it in other Apiaceae species. CtRLVaRNA therefore seems to play an important role in the determination of the degree of synergism or antagonism between the different members of the CMD complex and this interaction may also depend on the hosts species or possibly even on the host genotype (Yoshida et al. 2021). These findings parallel our results on co-occurrence of aphid-transmitted viruses (Chapter 3) with most of the positive associations detected between the dependent umbraviruses and associated subviral agents compared to fewer positive associations between the assistor poleroviruses or enamoviruses and the dependent agents. These results are also in keeping with those showing that the TVDVaRNA increases the accumulation of the umbravirus TBTV while the latter facilitated systemic movement of the associated RNA and the satellite RNA indicating a strong synergistic interaction between dependent viruses (Chen et al., 2022).

Even though satellites associated with helper-dependent diseases are known to play an important role in symptom expression such as in case of GRV (Naidu et al., 1998) or TBTD (Chen et al., 2022), the role of CMoVsatRNA is still not understood (Menzel et al., 2009). The situation is further complicated by the fact that umbraviruses and associated satellites can be promiscuous in their association with helper viruses, a finding that was shown first by experimental work (Barker, 1989; Mayo et al., 2000) and later through HTS sequencing, showing new combinations of novel and known poleroviruses and umbraviruses that are involved in helper dependent disease complexes such as TBTV (Abraham et al., 2019; Tan et al., 2021)

Within the present work key players of the CMD complex CtRLV, CMoV (and CMoMiV), and their associated subviral agents, CtRLVaRNA and CMoVsatRNA reached up to 100 % infection rates in plants sampled from cultivated fields. However, Nouvelle Aquitaine carrot growers do not seem concerned by viral diseases in general and by the high level of infection by viruses of the CMD complex, which is surprising given that CMD is reported as one of the most important carrot diseases (Stubbs, 1948; Watson and Serjeant, 1964; Krass and Schlegel, 1974; Adams et al., 2014; Yoshida, 2020). This raises the question of the reason(s) **why this extremely high prevalence of agents of a severe disease not perceived as a problem by**

growers? Are the obvious foliar yellowing symptoms misinterpreted as having other causes of no import? is there for some reason no CMD impact on production and yield? Several interesting hypotheses can be proposed and would be worth investigating. One is that the varieties used growers could somehow be tolerant to CMD and show limited impact on harvested roots. There is indeed evidence that varieties vary greatly with regard to their susceptibility to CMD (Stubbs 1948, Watson and Falk 1994) and that the outcome of the interaction of helper and dependent viruses might depend on the host genotype (Yoshida, 2020, Chen et al., 2022). Another hypothesis could be that additional unknown interaction between coinfecting agents might counterbalance the CMD negative impact. In this respect, it may be noteworthy that two recently described and poorly known agents, the wild carrot mottle umbravirus and the arracacha latent virus E-associated RNA were identified with a high prevalence in field carrots, raising the question of their potential impact on the subtle interplay between the various partners involved in the CMD complex. More detailed analyses contrasting symptoms on multiply infected plants with their virome and the titer reached by each agent or experimental inoculations with single or multiple agents could help to investigate these hypotheses. Understanding the reason(s) for the perceived or real limited impact of CMD in Nouvelle Aquitaine carrots could potentially bring novel ideas of sustainable solutions for the control of the many yellowing diseases caused by polero-, enamo- and luteoviruses currently re-emerging as a consequence of the ban of neonicotinoid insecticides formerly used to control these diseases and their vectors.

The results of HTS-based carrot virome analyses I performed during this thesis have provided deep insight into the complex and contrasted fingerprints of viral communities associated with field and wild carrot populations. The carrot pathosystem has proven particularly rich and original and the results obtained represent a baseline for further investigations on a range of fascinating questions including (i) the forces driving the assembly of these complex communities, (ii) the circulation of viruses at the agroecological interface and the unforeseen barriers that may limit this circulation and (iii) the complex interplay of interdependent viruses and virus-like agents functioning on a continuum between synergism and antagonism with important implications for their epidemiology and for the development of sustainable solutions in agriculture.

BIBLIOGRAPHY

Abbot P, Tooker J, Lawson SP (2018) Chemical ecology and sociality in aphids: opportunities and directions. *J. Chem. Ecol*. **44**, 770-784. https://doi.org/10.1007/s10886-018-0955-z

Abraham AD, Menzel W, Bekele B, Winter S (2014) A novel combination of a new umbravirus, a new satellite RNA and potato leafroll virus causes tobacco bushy top disease in Ethiopia. *Arch. Virol.* **159**, 3395-3399. https://doi.org/10.1007/s00705-014-2202-4

Adams IP, Glover RH, Monger WA, Mumford R, Jackeviciene E, Navalinskiene M, Samuitiene M, Boonham N (2009) Next-generation sequencing and metagenomic analysis: a universal diagnostic tool in plant virology. *Mol. Plant Pathol.* **10**, 537-545. https://doi.org/10.1111/j.1364-3703.2009.00545.x

Adams IP, Skelton A, Macarthur R, Hodges T, Hinds H, Flint L, Nath PD, Boonham N, Fox A (2014) Carrot yellow leaf virus is associated with carrot internal necrosis. *PLOS One*, **9**, e109125. https://doi.org/10.1371/journal.pone.0109125

Adams IP, Skelton A, Macarthur R, Hodges T, Hinds H, Flint L, Nath PD, Boonham N, Fox A (2014) Carrot yellow leaf virus is associated with carrot internal necrosis. *PLOS One* **9**, e109125. https://doi.org/10.1371/journal.pone.0109125

Afreen B, Khan AA, Naqvi QA, Kumar S, Pratap D, Snehi SK, Raj SK (2009) Molecular identification of a Cucumber mosaic virus subgroup II isolate from carrot (Daucus carota) based on RNA3 genome sequence analyses. *J. Plant. Dis. Protect.* **11**, 193-199. https://doi.org/10.1007/BF03356310

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. https://doi.org/10.1016/j.virol.2009.02.028

Alexander HM, Mauck KE, Whitfield AE, Garrett KA, Malmstrom CM (2014) Plant-virus interactions and the agro-ecological interface. *Eur. J. Plant Pathol.* **138**, 529-547. https://doi.org/10.1007/s10658-013-0317-1

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215,** 403-410. https://doi.org/10.1016/S0022-2836(05)80360-2

Athanasopoulou K, Boti MA, Adamopoulos PG, Skourou PC, Scorilas A (2022) Third-generation sequencing: the spearhead towards the radical transformation of modern genomics. *Life* **12**, 30. https://doi.org/10.3390/life12010030

Baker KK, Ramsdell DC, Gillett J (1985). Electron microscopy: Current applications to plant virology. *Plant dis.* **69**, 85-91. https://doi.org/10.1094/PD-69-85

Ball EM, Brakke MK (1968) Leaf-dip serology for electron microscopic identification of plant viruses. *Virology* **36,**152-55. https://doi.org/10.1016/0042-6822(68)90128-1

Baltimore D (1971) Viral genetic systems. *Trans. N.Y. Acad. Sci.* **33**, 327-332. https://doi.org/10.1111/j.2164-0947.1971.tb02600.x

Barker H (1989) Specificity of the effect of sap-transmissible viruses in increasing the accumulation of luteoviruses in co-infected plants. *Ann. Appl. Biol.* **115**, 71-78. https://doi.org/10.1111/j.1744-7348.1989.tb06813.x

Berend K, Haynes KR, MacKenzie CM (2019) Common garden experiments as a dynamic tool for ecological studies of alpine plants and communities in northeastern North America. *Rhodora*, **121**, 174-212. https://doi.org/10.3119/18-16

Bernardo P, Charles-Dominique T, Barakat M, Ortet P, Fernandez E, Filloux D, Hartnady P, Rebelo TA, Cousins SR, Mesleard F, Cohez D, Yavercovski N, Varsani A, Harkins GW, Peterschmitt M, Malmstrom CM, Martin DP, Roumagnac P (2018) Geometagenomics illuminates the impact of agriculture on the distribution and prevalence of plant viruses at the ecosystem scale. *ISME J.* **12**, 173-184. https://doi.org/10.1038/ismej.2017.155

Blawid R, Silva JMF, Nagata T (2017) Discovering and sequencing new plant viral genomes by next-generation sequencing: description of a practical pipeline. *Ann. Appl. Biol.* **170**, 301-314. https://doi.org/10.1111/aab.12345

Boccardo G, Lisa V, Luisoni E, Milne RG (1987) Cryptic plant viruses. *Adv. Virus Res.* **32**, 171-214. https://doi.org/10.1016/S0065-3527(08)60477-7

Boonham N, Tomlinson J, Mumford R (2007) Microarrays for rapid identification of plant viruses. *Ann. Rev. Phytopath.* **45**, 307-328. https://doi.org/10.1146/annurev.phyto.45.062806.094349

Bos L. (1981) Wild plants in the ecology of virus diseases. In Plant diseases and vectors: ecology and epidemiology, Maramarosch K, Harris KF (eds.), Academic press, NY (USA), pp 1-28. https://doi.org/10.1016/B978-0-12-470240-0.50005-6

Brakke MK (1953) Zonal separations by density-gradient centrifugation. *Arch. Biochem. Biophys.* **45**, 275-290. https://doi.org/10.1016/S0003-9861(53)80005-6

Breitbart M, Salamon P, Andresen B, Mahaffy JM, Segall AM, Mead D, Azam F, Rohwer F. (2002) Genomic analysis of uncultured marine viral communities. *Proc. Natl. Acad. Sci. USA* **99**, 14250-14255. https://doi.org/10.1073/pnas.202488399

Briard M, Le Clerc V, Peltier D, Suel A, Granger J, Dubois C, Geoffriau E (2006) Genetic resources management with networks: the example of the network "Carrot and other Daucus". Second International Symposium on Plant Genetic Resources of Horticultural Crops, Seoul, South Korea. https://hal.inrae.fr/hal-02757432

Brister JR, Ako-adjei D, Bao Y, Blinkova O (2015) NCBI Viral Genomes Resource. *Nucleic Acids Res.* **43**, D571-D577, https://doi.org/10.1093/nar/gku1207

Britten RJ, Davidson EH (1985) Hybridization strategy. In: Nucleic Acid Hybridization: A Practical Approach, Hames BD, Higgins SJ (Eds.), IRL Press, Oxford (UK), pp. 3-15.

Broadbent LH (1950) The correlation of aphid numbers with the spread of leaf roll and rugose mosaic in potato crops. *Ann. Appl. Biol.* **37**, 58-65. https://doi.org/10.1111/j.1744-7348.1950.tb00950.x

Brunt AA, Crabtree K, Dallwitz MJ, Gibbs AJ, Watson L, Zurcher EJ (eds.) (1996). Plant Viruses Online: Descriptions and Lists from the VIDE Database. Version: 20th August 1996. http://biology.anu.edu.au/Groups/MES/vide/

Buchfink B, Xie C, Huson DH (2014) Fast and sensitive protein alignment using Diamond. *Nat. Methods* **12**, 59-60. https://doi.org/10.1038/nmeth.3176

Campbell AJ, Erickson A, Pellerin E, Salem N, Mo X, Falk BW, Ferriol I (2020) Phylogenetic classification of a group of self-replicating RNAs that are common in co-infections with poleroviruses. *Virus Res.* **276**, 197831. https://doi.org/10.1016/j.virusres.2019.197831

Candresse T, Fuchs M (2020) *Closteroviridae*. Encyclopedia of Life Sciences, *eLS* **2020**, 1-10. https://doi.org/10.1002/9780470015902.a0000747.pub4

Carmo-Sousa M, Moreno A, Garzo E, Fereres A (2014) A non-persistently transmitted-virus induces a pull–push strategy in its aphid vector to optimize transmission and spread. *Virus Res.* **186**, 38-46. https://doi.org/https://doi.org/10.1016/j.virusres.2013.12.012

Chen X, Luo H, Zhang J, Ma Y, Li K, Xiong F, Yang Y, Yang J, Lan P, Wei T, Xu Y, Chen H, Li F (2022) Synergism Among the Four Tobacco Bushy Top Disease Casual Agents in Symptom Induction and Aphid Transmission. *Frontiers in Microbiology*, 13. https://doi.org/10.3389/fmicb.2022.846857

Civitello DJ, Cohen J, Fatima H, Halstead NT, Liriano J, McMahon TA, Ortega CN, Sauer EL, Sehgal T, Young S, Rohr JR (2015) Biodiversity inhibits parasites: broad evidence for the dilution effect. *Proc. Natl. Acad Sci. USA* **112**, 8667-8671. https://doi.org/10.1073/pnas.1506279112

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

Cooper I, Jones RAC (2006) Wild plants and viruses: under-investigated ecosystems. *Adv. Virus Res.* **67,** 1-47. https://doi.org/10.1016/S0065-3527(06)67001-2

Csorba T, Kontra L, Burgyán J (2015) Viral silencing suppressors: tools forged to fine-tune host-pathogen coexistence. *Virology* **479-480**, 85-103. https://doi.org/10.1016/j.virol.2015.02.028

Deamer D, Akeson M, Branton D (2016) Three decades of nanopore sequencing. *Nat. Biotechnol*. **34**, 518-524. https://doi.org/10.1038/nbt.3423

Dodds JA, Morris TJ, Jordan RL (1984) Plant viral double-stranded RNA. *Ann. Rev. Phytopath.* **22**, 151-168. https://doi.org/10.1146/annurev.py.22.090184.001055

Donaire L, Wang Y, Gonzalez-Ibeas D, Mayer KF, Aranda MA, Llave C (2009) Deep-sequencing of plant viral small RNAs reveals effective and widespread targeting of viral genomes. *Virology* **392**, 203-214. https://doi.org/10.1016/j.virol.2009.07.005

Duffus JE, Larsen RC, Liu HY (1986) Lettuce infectious yellows virus - a new type of whitefly-transmitted virus. *Phytopathology* ***76,*** 97-100. https://doi.org/10.1094/Phyto-76-97

Dunn JA (1965) Studies on the aphid, *Cavariella aegopodii* Scop. *Ann. Appl. Biol.* **56**, 429-438. https://doi.org/10.1111/j.1744-7348.1966.tb04380.x

Elena SF, Fraile A, García-Arenal F (2014) Evolution and emergence of plant viruses. *Adv. Virus Res.* **88**, 161-191. https://doi.org/10.1016/B978-0-12-800098-4.00003-9

Elnagar S, Murant AF (1978) Aphid-injection experiments with carrot mottle virus and its helper virus, carrot red leaf. *Ann. Appl. Biol*. **89**, 245-250. https://doi.org/10.1111/j.1744-7348.1978.tb07696.x

Fagoaga C, Duran-Vila N (1996) Naturally occurring variants of citrus exocortis viroid in vegetable crops. *Plant Pathol.* **45**, 45-53. https://doi.org/10.1046/j.1365-3059.1996.d01-104.x

Fargette D, Konaté G, Fauquet C, Muller E, Peterschmitt M, Thresh JM (2006) Molecular ecology and emergence of tropical plant viruses, *Ann. Rev. Phytopath.* **44**, 235-260. https://doi.org/10.1146/annurev.phyto.44.120705.104644

Fetters AM, Cantalupo PG, Wei N, Robles MTS, Stanley A, Stephens JD, Pipas JM, Ashman TL (2022) The pollen virome of wild plants and its association with variation in floral traits and land use. *Nat. Comm.* **13**, 523. https://doi.org/10.1038/s41467-022-28143-9

Fraenkel-Conrat H (1986) The history of tobacco mosaic virus and the evolution of molecular biology. In The Plant Viruses, van Regenmortel MHV, Fraenkel-Conrat H (Eds), Springer, Boston (USA), pp 5-17. https://doi.org/10.1007/978-1-4684-7026-0_1

Fraile A, García-Arenal F (2016) Environment and evolution modulate plant virus pathogenesis. *Curr. Opin. Virol.* **17**, 50-56. https://doi.org/10.1016/j.coviro.2016.01.008

Fraile A, McLeish MJ, Pagán I, González-Jara P, Piñero D, García-Arenal F (2017) Environmental heterogeneity and the evolution of plant-virus interactions: viruses in wild pepper populations. *Virus Res.* **241**, 68-76. https://doi.org/10.1016/j.virusres.2017.05.015

François S, Filloux D, Frayssinet M, Roumagnac P, Martin DP, Ogliastro M, Froissart R (2018) Increase in taxonomic assignment efficiency of viral reads in metagenomic studies. *Virus Res.* **244**, 230-234. https://doi.org/10.1016/j.virusres.2017.11.011

Galan M, Razzauti M, Bard E, Bernard M, Brouat C, Charbonnel N, Dehne-Garcia A, Loiseau A, Tatard C, Tamisier L, Vayssier-Taussat M, Vignes H, Cosson JF (2016) 16S rRNA Amplicon Sequencing for Epidemiological Surveys of Bacteria in Wildlife. *Msystems* **1**, e00032-16. https://doi.org/10.1128/mSystems.00032-16

Galibert F, Mandart E, Fitoussi F, Tiollais P, Charnay P (1979) Nucleotide sequence of the hepatitis B virus genome (subtype ayw) cloned in *E. coli*. *Nature* **281**, 646-650. https://doi.org/10.1038/281646a0

Geoffriau E, Simon PW (2020). Carrots and related *Apiaceae* crops, 2nd edition. CABI Publishing, Wallingford (UK), 368 pp. https://doi.org/10.1079/9781789240955.0000

Gibbs AJ (1983) Virus ecology - struggle of the genes. In Physiological Plant Ecology III. Encyclopedia of Plant Physiology, Lange OL, Nobel PS, Osmond CB, Zieger H (eds) Springer, Berlin (Germany), pp. 537-558.

Gibbs MJ, Cooper JI, Waterhouse PM (1996) The genome organization and affinities of an Australian isolate of carrot mottle umbravirus. *Virology* **224**, 310-313. https://doi.org/10.1006/viro.1996.0533

Gorbalenya AE, Krupovic M, Mushegian A, Kropinski AM, Siddell SG, Varsani A, Adams MJ, Davison AJ, Dutilh BE, Harrach B, Harrison RL, Junglen S, King AMQ, Knowles NJ, Lefkowitz, EJ, Nibert ML, Rubino L, Sabanadzovic S, Sanfaçon H, Simmonds P, Walker PJ, Zerbini FM, Kuhn, JH (2020) The new scope of virus taxonomy: partitioning the virosphere into 15 hierarchical ranks. *Nat. Microbiol.* **5**, 668-674. https://doi.org/10.1038/s41564-020-0709-x

Gorovits R, Sobol I, Altaleb M, Czosnek H, Anfoka G (2019) Taking advantage of a pathogen: understanding how a virus alleviates plant stress response. *Phytopathol. Res.* **1**, 20. https://doi.org/10.1186/s42483-019-0028-4

Groen SC, Jiang S, Murphy AM, Cunniffe NJ, Westwood JH, Matthew PD, Bruce TJA, Caulifield JC, Furzer OJ, Reed A, Robinson SI, Miller E, Davis CN, Pickett JA, Whitney HM, Glover, BJ, Carr JP (2016) Virus infection of plants alters pollinator preference: a payback for susceptible hosts? *PLOS Pathog.* **12**, e1005790. https://doi.org/10.1371/journal.ppat.1005790

Hadidi A, Candresse T (2003) Polymerase chain reaction. In: Hadidi A, Flores R, Randles JW, Semancik JS (Eds.) Viroids. CSIRO Publishing, Collingwood, VIC, pp. 115-122.

Handelsman J (2005) Metagenomics: application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.* **69,** 195-195. https://doi.org/10.1128/MMBR.68.4.669-685.2004

Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem. Biol.* **5**, R245-R249. https://doi.org/10.1016/S1074-5521(98)90108-9

Hasiów-Jaroszewska B, Boezen D, Zwart MP (2021) Metagenomic studies of viruses in weeds and wild plants: a powerful approach to characterise variable virus communities. *Viruses* **13**, 1939. https://doi.org/10.3390/v13101939

Heather JM, Chain B (2016) The sequence of sequencers: The history of sequencing DNA. *Genomics* **107**, 1-8. https://doi.org/10.1016/j.ygeno.2015.11.003

Henson JM, French R (1993) The Polymerase Chain Reaction and Plant Disease Diagnosis. *Ann. Rev. Phytopathol.* **31**, 81-109. https://doi.org/10.1146/annurev.py.31.090193.000501

Hogenhout SA, Ammar ED, Whitfield AE, Redinbaugh MG (2008) Insect vector interactions with persistently transmitted viruses. *Ann. Rev. Phytopath.* **46**, 327-359. https://doi.org/10.1146/annurev.phyto.022508.092135

Howell WE and Mink GI (1976) Host range, purification, and properties of a flexuous rod-shaped virus isolated from carrot. *Phytopathology* **66,** 949-953. https://www.apsnet.org/publications/phytopathology/backissues/Documents/1976Articles/phyto66n08_949.PDF

Howell WE, Mink GI (1981) Viruses isolated from wild carrot and poison hemlock. *Plant Dis.* **65**, 277-279. https://doi.org/10.1094/PD-65-277

Hull R (2014) Plant virology. Fifth edition. Academic Press, San Diego (USA). https://doi.org/10.1016/B978-0-12-384871-0.00001-7

Hull R, Adams AN (1968) Groundnut rosette and its assistor virus. *Ann. Appl. Biol.* **62**, 139-145. https://doi.org/10.1111/j.1744-7348.1968.tb03857.x

Illumina (2017) Effects of index misassignment on multiplexing and downstream analysis. https://www.illumina.com/content/dam/illumina-marketing/documents/products/whitepapers/index-hopping-white-paper-770-2017-004.pdf

Iorizzo M, Senalik DA, Ellison SL, Grzebelus D, Cavagnaro PF, Allender C, Brunet J, Spooner DM, Van Deynze A, Simon PW (2013) Genetic structure and domestication of carrot (*Daucus carota* subsp. sativus) (Apiaceae). *Am. J. Bot.* **100**, 930-938. https://doi.org/10.3732/ajb.1300055

Koskela T (2002) Variation in life-history traits among *Urtica dioica* populations with different history in parasitism by the holoparasitic plant *Cuscuta europaea*. *Evol. Ecol.* **16**, 433-454. https://doi.org/10.1023/A:1020860718125

Krass CJ, Schlegel DE (1974) 'Motley dwarf' virus disease complex of california carrots. *Phytopathology* **64**, 151-152. https://www.apsnet.org/publications/phytopathology/backissues/Documents/1974Articles/Phyto64n01_151.PDF

Kreuze JF, 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. https://doi.org/10.1016/j.virol.2009.03.024

Kumar J, Gunapati S, Singh SP, Gadre R, Sharma NC, Tuli R (2013) Molecular characterization and pathogenicity of a carrot (Daucus carota) infecting begomovirus and associated betasatellite from India. *Virus Res.* **178**, 478-485. https://doi.org/10.1016/j.virusres.2013.10.010

Kundu R and Dixon A (1993) Do host alternating aphids know which plant they are on? *Ecol. Entomol.*, **18**, 61-66. https://doi.org/10.1111/j.1365-2311.1993.tb01080.x

Kutnjak D, Tamisier L, Adams I, Boonham N, Candresse T, Chiumenti M, De Jonghe K, Kreuze JF, Lefebvre M, Silva G, Malapi-Wight M, Margaria P, Mavrič Pleško I, McGreig S, Miozzi L, Remenant B, Reynard JS, 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. https://doi.org/10.3390/microorganisms9040841

Laine AL (2007) Detecting local adaptation in a natural plant–pathogen metapopulation: a laboratory vs. field transplant approach. *J. Evol. Biol.* **20**, 1665-1673. https://doi.org/10.1111/j.1420-9101.2007.01359.x

Lebouvier M, Laparie M, Hulle M, Marais A, Cozic Y, Lalouette L, Vernon P, Candresse T, Frenot Y, Renault D (2011). The significance of the sub-Antarctic Kerguelen Islands for the assessment of the vulnerability of native communities to climate change, alien insect invasions and plant viruses. *Biol. Invasions*, **13**, 1195-1208. https://doi.org/10.1007/s10530-011-9946-5

Legg JP (1999) Emergence, spread and strategies for controlling the pandemic of cassava mosaic virus disease in east and central Africa. *Crop Protect.* **18**, 627-637. https://doi.org/10.1016/S0261-2194(99)00062-9

Liefting LW, Waite DW, Thompson JR (2021) Application of Oxford Nanopore technology to plant virus detection. *Viruses* **13**, 1424. https://doi.org/10.3390/v13081424

Lin NS, Hsu YH, Hsu HT (1990) Immunological detection of plant viruses and a mycoplasmalike organism by direct tissue blotting on nitrocellulose membranes. *Phytopathology* **80**, 824-828. https://doi.org/10.1094/Phyto-80-824

Linke B, Alessandro MS, Galmarini CR, Nothnagel T (2019) Carrot floral development and reproductive biology. In: The Carrot Genome. Compendium of Plant Genomes, Simon P, Iorizzo M, Grzebelus D, Baranski, R (Eds), Springer, Cham (Switzerland), pp 27-57. https://doi.org/10.1007/978-3-030-03389-7_3

Liu X, Chen L, Liu M, Garcıa-Guzman G, Gilbert GS, Zhou S (2020) Dilution effect of plant diversity on infectious diseases: latitudinal trend and biological context dependence. *Oikos* **129**, 457-465. https://doi.org/10.1111/oik.07027

Lovisolo O, Hull R, Rosler O (2003) Coevolution of viruses with hosts and vectors and possible paleontology. *Adv. Virus Res.* **62**, 325-379. https://doi.org/10.1016/S0065-3527(03)62006-3

Ma Y, Fort T, Marais A, Lefebvre M, Theil S, Vacher C, Candresse T (2021) Leaf-associated fungal and viral communities of wild plant populations differ between cultivated and natural ecosystems, *Plant-Environ. Interact.* **2**, 87-99. https://doi.org/10.1002/pei3.10043

Ma Y, Marais A, Lefebvre M, Faure C, Candresse T (2020) Metagenomic analysis of virome cross-talk between cultivated Solanum lycopersicum and wild *Solanum nigrum*. *Virology* **540**, 38-44. https://doi.org/10.1016/j.virol.2019.11.009

Maachi A, Donaire L, Hernando Y, Aranda MA (2022) Genetic differentiation and migration fluxes of viruses from melon crops and crop edge weeds. *J. Virol*. **96**, e0042122. https://doi.org/10.1128/jvi.00421-22

Maclot F, Candresse T, Filloux D, Malmstrom CM, Roumagnac P, van der Vlugt R, Massart S (2020) Illuminating an Ecological Blackbox: Using High Throughput Sequencing to Characterize the Plant Virome Across Scales. *Front Microbiol*. **11**, 578064. https://doi.org/10.3389/fmicb.2020.578064

Malmstrom CM, Alexander HM (2016) Effects of crop viruses on wild plants. *Curr. Opin. Virol.* **19**, 30-36. https://doi.org/10.1016/j.coviro.2016.06.008

Malmstrom CM, McCullough AJ, Johnson HA, Newton LA, Borer IT (2005) Invasive annual grasses indirectly increase virus incidence in California native perennial bunchgrasses, *Oecologia*, **145**, 153-164. https://doi.org/10.1007/s00442-005-0099-z

Malmstrom CM, Melcher U, Bosque-Pérez NA (2011) The expanding field of plant virus ecology: Historical foundations, knowledge gaps, and research directions. *Virus Res.* **159**, 84-94. https://doi.org/10.1016/j.virusres.2011.05.010

Malyshev AV, Henry HAL, Kreyling J (2014) Common garden experiments to characterize cold acclimation responses in plants from different climatic regions. In: Plant cold acclimation: methods and protocols, Hincha DK, Zuther E (Eds.), Springer, New York (USA), pp 65-78. https://doi.org/10.1007/978-1-4939-0844-8_7

Mandel JR, Brunet J (2019) Gene flow in carrot. In The Carrot Genome. Compendium of Plant Genomes. Simon P, Iorizzo M, Grzebelus D, Baranski, R (Eds), Springer, Cham (Switzerland), pp 59-76. https://doi.org/10.1007/978-3-030-03389-7_4

Maree HJ, Fox A, Al Rwahnih M, Boonham N, Candresse T (2018) Application of hts for routine plant virus diagnostics: state of the art and challenges. *Front. Plant Sci.* **9**, 1-4. https://doi.org/10.3389/fpls.2018.01082

Mascia T, Gallitelli D (2016) Synergies and antagonisms in virus interactions. *Plant Sci.* **252**, 176-192. https://doi.org/10.1016/j.plantsci.2016.07.015

Maskell LC, Raybould AF, Cooper JI, Edwards ML, Gray AJ (1999) Effects of turnip mosaic virus and turnip yellow mosaic virus on the survival, growth and reproduction of wild cabbage (Brassica oleracea). *Ann. Appl. Biol.* **135,** 401-407. https://doi.org/10.1111/j.1744-7348.1999.tb00867.x

Massart S, Adams I, Al Rwahnih M, Baeyen S, Bilodeau GJ, Blouin AG, Boonham N, Candresse T, Chandelier A, de Jonghe K, Fox A, Gaafar YZA, 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, Martoni F, Mehle N, Minafra A, Mollov D, Moreira A, Nakhla M, Petter F, Piper AM, Ponchart J, Rae R, Remenant B, Rivera Y, Rodoni B, Botermans M, Roenhorst JW, Rollin J, Saldarelli P, Santala J, Souza-Richards R, Spadaro D, Studholme DJ, Sultmanis S, van der Vlugt R, Tamisier L, Trontin C, Vazquez I, Vicente CSL, Vosseberg BTLH, Westenberg M, Wetzel T, Ziebell H, Lebas BS (2022). *Peer Community J*. **2**, e62. https://doi.org/10.24072/pcjournal.181

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

Matthews REF (1991) Plant Virology (Third Edition), Academic Press, San Diego (USA).

Mauck K, Bosque-Pérez NA, Eigenbrode SD, De Moraes CM, Mescher MC (2012) Transmission mechanisms shape pathogen effects on host-vector interactions: Evidence from plant viruses. *Funct. Ecol*. **26,** 1162-1175. https://doi.org/10.1111/j.1365-2435.2012.02026.x

Mauck KE, De Moraes CM, Mescher MC (2010) Deceptive chemical signals induced by a plant virus attract insect vectors to inferior hosts. *Proc. Natl. Acad. Sci. USA* **107**, 3600-3605. https://doi.org/10.1073/pnas.0907191107

Mayo M, Ryabov E, Fraser G, Taliansky M (2000) Mechanical transmission of potato leafroll virus. *J. Gen. Virol.* **81**, 2791-2795. https://doi.org/10.1099/0022-1317-81-11-2791

Melcher U, Muthukumar V, Wiley GB, Min BE, Palmer MW, Verchot-Lubicz J, Ali A, Nelson RS, Roe BA, Thapa V, Pierce ML (2008) Evidence for novel viruses by analysis of nucleic acids in virus-like particle fractions from *Ambrosia psilostachya*. *J. Virol. Meth.* **152**, 49-55. https://doi.org/10.1016/j.jviromet.2008.05.030

Menzel W, Maiss E, Vetten HJ (2009) Nucleotide sequence of a satellite RNA associated with carrot motley dwarf in parsley and carrot. *Virus Genes*, **38**, 187-188. https://doi.org/10.1007/s11262-008-0302-5

Menzel W, Vetten HJ (2008a) Partial nucleotide sequence of a carlavirus from carrot (*Daucus carota* L.). *Virus Genes* **37**, 432-433. https://doi.org/10.1007/s11262-008-0276-3

Menzel W, Vetten HJ (2008b) Complete nucleotide sequence of an isolate of the Anthriscus strain of Parsnip yellow fleck virus. *Arch. Virol.* **153**, 2173-2175. https://doi.org/10.1007/s00705-008-0240-5

Miclotte G, Heydari M, Demeester P, Rombauts S, Van de Peer Y, Audenaert P, Fostier J (2016) Jabba: hybrid error correction for long sequencing reads. *Algorithms Mol. Biol.* **11**, 10. https://doi.org/10.1186/s13015-016-0075-7

Mihara T, Nishimura Y, Shimizu Y, Nishiyama H, Yoshikawa G, Uehara H, Hingamp P, Goto S, Ogata H (2016) Linking virus genomes with host taxonomy. *Viruses* **8**, 66. https://doi.org/10.3390/v8030066

Mo XH, Chen ZB, Chen JP (2011) Molecular identification and phylogenetic analysis of a viral RNA associated with the Chinese tobacco bushy top disease complex. *Ann. Appl. Biol.* **158**, 188-193. https://doi.org/10.1111/j.1744-7348.2010.00452.x

Moloney KA, Holzapfel C, Tielbörger K, Jeltsch F, Schurr FM (2009) Rethinking the common garden in invasion research. *Perspect. Plant Ecol. Evol. Syst.* **11**, 311-320. https://doi.org/10.1016/j.ppees.2009.05.002

Moran J, van Rijswijk B, Traicevski V, Kitajima EW, Mackenzie AM, Gibbs AJ (2002) Potyviruses, novel and known, in cultivated and wild species of the family Apiaceae in Australia. *Arch. Virol.* **147**, 1855-1867. https://doi.org/10.1007/s00705-002-0865-8

Moubset O, François S, Maclot F, Palanga E, Julian C, Claude L, Fernandez E, Rott PCE, Daugrois JH, Antoine-Lorquin A, Bernardo P, Blouin AG, Temple C, Kraberger S, Fontenele RS, Harkins GW, Marais A, Candresse T, Ben Chéhida S, Lefeuvre P, Lett JM, Varsani A, Massart S, Ogliastro MH, Martin DP, Filloux D, Roumagnac P (2022) Virion-

associated nucleic acid-based metagenomics: a decade of advances in molecular characterization of plant viruses. *Phytopathology*, Online first. https://doi.org/10.1094/PHYTO-03-22-0096-RVW

Moury B, Fabre F, Hébrard E, Froissart R (2017) Determinants of host species range in plant viruses. *J. Gen. Virol.* **98**, 862-873. https://doi.org/10.1099/jgv.0.000742

Murant AF (1975) Occurrence of mottle and redleaf components of carrot motley dwarf disease in British Columbia. *Can. Plant Dis. Surv.* **55**, 103-105. https://phytopath.ca/wp-content/uploads/2014/10/cpds-archive/vol55/CPDS_Vol_55_No_3_(103-105)1975.pdf

Muthukumar V, Melcher U, Pierce M, Wiley GB, Roe BA, Palmer MW, Thapa V, Ali A, Ding T (2009) Non-cultivated plants of the Tallgrass Prairie Preserve of northeastern Oklahoma frequently contain virus-like sequences in particulate fractions. *Virus Res.* **141**, 169-173. https://doi.org/10.1016/j.virusres.2008.06.016

Naidu RA, Bottenberg H, Subrahmanyam P, Kimmins FM, Robinson DJ, Thresh JM (1998) Epidemiology of groundnut rosette virus disease: current status and future research needs. *Ann. Appl. Biol.* **132**, 525-548. https://doi.org/10.1111/j.1744-7348.1998.tb05227.x

Nakatsukasa-Akune M, Yamashita K, Shimoda Y, Uchiumi T, Abe M, Aoki T, Kamizawa A, Ayabe S, Higashi S, Suzuki A (2005) Suppression of root nodule formation by artificial expression of the TrEnodDR1 (coat protein of White clover cryptic virus 2) gene in *Lotus japonicus*. *Molec. Plant-Microb. Interact.* **18**, 1069-1080. https://doi.org/10.1094/MPMI-18-1069

Naseem MT, Ashfaq M, Khan, AM, Kiss Z, Akhtar KP, Mansoor S (2016) Transmission of viruses associated with carrot motley dwarf by *Myzus persicae*. *J. Plant Pathol.* **98**, 581-585. https://doi.org/10.4454/JPP.V98I3.026

Natsuaki T, Muroi Y, Okuda S, Teranaka M (1990) Carrot temperate cryptoviruses and RNA-RNA hybridization among their double-stranded RNA segments. *Ann. Phytopath. Soc. Japan* **56**, 354-358. https://www.jstage.jst.go.jp/article/jjphytopath1918/56/3/56_3_354/_pdf

Pagán I, González-Jara P, Moreno-Letelier A, Rodelo-Urrego M, Fraile A, Piñero D, García-Arenal F. (2012) Effect of biodiversity changes in disease risk: exploring disease emergence in a plant-virus system. *PLOS Pathogens* **8**, e1002796. https://doi.org/10.1371/journal.ppat.1002796

Parry JN, PersleyDM (2005) Carrot as a natural host of Watermelon mosaic virus. Australian *Plant Pathol.* **34**, 283-284. https://doi.org/10.1071/AP05016

Petruzzello M (2019) List of plants in the family *Apiaceae*. *Encyclopedia Britannica.* https://www.britannica.com/topic/list-of-plants-in-the-family-Apiaceae-2038061. Accessed 30 October 2022

Pruss G, Ge X, Shi XM, Carrington JC and Vance VB (1997) Plant viral synergism: the potyviral genome encodes a broad-range pathogenicity enhancer that transactivates replication of heterologous viruses. *Plant Cell* **9**, 859-868. https://doi.org/10.1105/tpc.9.6.859

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD (2012) The Pfam protein families database. *Nucl. Acids Res.* **40**, D290-D301. https://doi.org/10.1093/nar/gkr1065

Que F, Hou XL, Wang GL, Xu ZS, Tan GF, Li T, Wang YH, Khadr A, Xiong AS (2019) Advances in research on the carrot, an important root vegetable in the *Apiaceae* family. *Hortic. Res.* **6**, 69. https://doi.org/10.1038/s41438-019-0150-6

Rhymer JM (1992) An experimental study of geographic variation in avian growth and development. *J. Evol. Biol.* **5**, 289-306. https://doi.org/10.1046/j.1420-9101.1992.5020289.x

Robinson DJ, Ryabov EV, Raj SK, Roberts IM. Taliansky ME (1999) Satellite RNA is essential for encapsidation of groundnut rosette umbravirus RNA by groundnut rosette assistor luteovirus coat protein. *Virology* **254**, 105-114. https://doi.org/10.1006/viro.1998.9527

Ronfort J, Saumitou-Laprade P, Cugen J, Couvet D (1995) Mitochondrial DNA diversity and male sterility in natural populations of *Daucus carota* ssp carota. *Theor. Appl. Genet.* **91**, 150-159. https://doi.org/10.1007/BF00220872

Rong J, Lammers Y, Strasburg JL, Schidlo NS, Ariyurek Y, de Jong TJ, Klinkhamer PGL, Smulders MJM, Vrieling K (2014) New insights into domestication of carrot from root transcriptome analyses. *BMC Genomics* **15**, 1-15. http://www.biomedcentral.com/1471-2164/15/895

Roossinck MJ (2012a) Persistent plant viruses: molecular hitchhikers or epigenetic elements. In Viruses: essential agents of life, Witzany G (Ed), Springer, New York (USA), pp. 177-186.

Roossinck MJ (2012b) Plant virus metagenomics: biodiversity and ecology. *Ann. Rev. Genet.* **46,** 359-369. https://doi.org/10.1146/annurev-genet-110711-155600

Roossinck MJ (2015b) Plants, viruses and the environment: ecology and mutualism. *Virology* **480**, 271-277. https://doi.org/10.1016/j.virol.2015.03.041

Roossinck MJ, García-Arenal F (2015) Ecosystem simplification, biodiversity loss and plant virus emergence. *Curr. Opin. Virol.* **10**, 56-62. https://dx.doi.org/10.1016/j.coviro.2015.01.005

Roossinck MJ, Martin DP, Roumagnac P (2015) Plant virus metagenomics: advances in virus discovery. *Phytopathology* **105**, 716-727. https://doi.org/10.1094/PHYTO-12-14-0356-RVW

Roossinck MJ, Saha P, Wiley GB, Quan J, White JD, Lai H, Chavarría F, Shen G, Roe BA (2010) Ecogenomics: using massively parallel pyrosequencing to understand virus ecology. *Mol. Ecol.* **19**, 81-88. https://doi.org/10.1111/j.1365-294X.2009.04470.x

Roossinck MJ. (2015a) Metagenomics of plant and fungal viruses reveals an abundance of persistent lifestyles. *Front. Microbiol*. **5**, 767. https://doi.org/10.3389/fmicb.2014.00767

Rozado-Aguirre Z, Adams I, Collins L, Fox A, Dickinson M, Boonham N (2016) Detection and transmission of Carrot torrado virus, a novel putative member of the Torradovirus genus. *J. Virol. Meth.* **235**,119-124. https://doi.org/10.1016/j.jviromet.2016.05.018

Ryabov EV, Robinson DJ, Taliansky M (2001) Umbravirus-encoded proteins both stabilize heterologous viral RNA and mediate its systemic movement in some plant species. *Virology*, **288**, 391-400. https://doi.org/10.1006/viro.2001.1078

Ryabov EV, Robinson DJ, Taliansky ME (1999) A plant virus-encoded protein facilitates long-distance movement of heterologous viral RNA. *Proc. Natl. Acad. Sci. USA*, **96**, 1212-1217. https://doi.org/10.1073/pnas.96.4.1212

Sallinen S, Norberg A, Susi H, Laine AL (2020) Intraspecific host variation plays a key role in virus community assembly. *Nat. Comm.* **11**, 5610. https://doi.org/10.1038/s41467-020-19273-z

Sanger F, Coulson AR (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* **94**, 441-448. https://doi.org/10.1016/0022-2836(75)90213-2

Saritha RK, Shrawan S, Kalia P, Jain RK, Baranwal VK (2017) Association of Pedilanthus leaf curl virus and satellites with carrot (Daucus carota) in India. *Plant Dis* **101**, 843. https://doi.org/10.1094/PDIS-05-16-0672-PDN

Sasu MA, Ferrari MJ, Du D, Winsor JA, Stephenson AG (2009) Indirect costs of a nontarget pathogen mitigate the direct benefits of a virus-resistant transgene in wild *Cucurbita*. *Proc. Natl. Acad. Sci. USA*, **106**, 19067-19071. https://doi.org/10.1073/pnas.0905106106

Saunders K, Bedford ID, Yahara T, Stanley J (2003) The earliest recorded plant virus disease. *Nature* **422**, 831. https://doi.org/10.1038/422831a

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467-70. https://doi.org/10.1126/science.270.5235.467

Siddell SG, Walker PJ, Lefkowitz EJ, Mushegian AR, Dutilh BE, Harrach B, Harrison RL, Junglen S, Knowles NJ, Kropinski AM, Krupovic M, Kuhn JH, Nilbert ML, Rubino L, Sabanadzovic S, Simmonds P, Varsani A, Zerbini FM, Davison AJ (2020) Binomial nomenclature for virus species: a consultation. *Arch. Virol.* **165**, 519-525. https://doi.org/10.1007/s00705-019-04477-6

Simmonds P, Adams MJ, Benkő M, Breitbart M, Brister JR, Carstens EB, Davison AJ, Delwart E, Gorbalenya AE, Harrach B, Hull R, King AMQ, Koonin EV, Krupovic M, Kuhn JH, Lefkowitz, EJ, Nibert ML, Orton R, Roossinck MJ, Sabanadzovic S, Sullivan MB, Suttle CA, Tesh RB, van der Vlugt, RA, Varsani A, Zerbini FM (2017) Consensus statement: virus taxonomy in the age of metagenomics. *Nat. Rev. Microbiol*. **15**, 161-168. https://doi.org/10.1038/nrmicro.2016.177

Sivalingam PN, Sumiya KV, Malathi VG (2011) Carrot as a new host for a begomovirus: yellow mosaic disease of carrot reported in India. *New Dis. Rep.* **23**, 34. https://doi.org/10.5197/j.2044-0588.2011.023.034

Soliman AM, Ahmed AA, Mohamed MS Hanna DH, Barsoum BN (2012) Gazar virus Y, a new member of the celery mosaic virus group of potyviruses, isolated from carrots in Egypt. *Australasian Plant Pathol.* **41**, 529-534. https://doi.org/10.1007/s13313-012-0148-0

Spurr CJ, Lucas AM (2021) *Apiaceae* seed production. In: Carrots and Related Apiaceae Crops, 2nd edn, Geoffriau E, Simon PW (Eds), CABI Publishing, Wallingford (UK), pp 18-34.

Stubbs LL (1948) A new virus disease of carrots: Its transmission, host range, and control. *Austr. J. Biol. Sci.* **1**, 303-332. https://doi.org/10.1071/BI9480303

Studholme DJ, Glover RH, Boonham N (2011) Application of high-throughput DNA sequencing in phytopathology. *Ann. Rev. Phytopath.* ***49,*** 87-105. https://doi.org/10.1146/annurev-phyto-072910-095408

Susi H, Filloux D, Frilander MJ, Roumagnac P, Laine AL (2019) Diverse and variable virus communities in wild plant populations revealed by metagenomic tools. *Peer J.* **7**, e6140. https://doi.org/10.7717/peerj.6140

Syller J (2003) Molecular and biological features of umbraviruses, the unusual plant viruses lacking genetic information for a capsid protein. *Physiol. Mol. Plant Pathol.* **63**, 35-46. https://doi.org/10.1016/j.pmpp.2003.08.004

Tan ST, Liu F, Lv J, Liu QL, Luo HM, Xu Y, Ma Y, Chen XJ, Lan PX, Chen HR, Cao MJ, Li F (2021) Identification of two novel poleroviruses and the occurrence of Tobacco bushy top disease causal agents in natural plants. *Sci. Rep.* **26**, 21045. https://doi: 10.1038/s41598-021-99320-x

Thapa V, McGlinn DJ, Melcher U, Palmer MW, Roossinck MJ (2015) Determinants of taxonomic composition of plant viruses at the Nature Conservancy's Tallgrass Prairie Preserve, Oklahoma. *Virus Evol*. **1**, 7. https://doi.org/10.1093/ve/vev007

Thompson JD, McNeilly T, Gay AJ (1991) Population variation in *Spartina anglica* C. E. Hubbard. *New Phytol.* **117**, 115-128. https://doi.org/10.1111/j.1469-8137.1991.tb00951.x

Valverde RA, Gutierrez DL (2007) Transmission of a dsRNA in bell pepper and evidence that it consists of the genome of an endornavirus. *Virus Genes* **35**, 399-403. https://doi.org/10.1007/s11262-007-0092-1

van Dijk P, Bos L (1985) Viral dieback of carrot and other *Umbelliferae* caused by the *Anthriscus* strain of parsnip yellow fleck virus, and its distinction from carrot motley dwarf. *Nether. J. Plant Pathol*. **91**, 169-187. https://doi.org/10.1007/BF02009678

Villamor DEV, Ho T, Al Rwahnih M, Martin RR, Tzanetakis IE (2019). High throughput sequencing for plant virus detection and discovery. *Phytopathology* **109**, 716-725. https://doi.org/10.1094/PHYTO-07-18-0257-RVW

Wang D, Coscoy L, Zylberberg M, Avila PC, Boushey HA, Ganem D, DeRisi JL (2002) Microarray-based detection and genotyping of viral pathogens. *Proc. Natl. Acad. Sci. USA* **99**, 15687-1569. https://doi.org/10.1073/pnas.242579699

Waterhouse PM, Murant AF (1983) Further evidence on the nature of the dependence of carrot mottle virus on carrot red leaf virus for transmission by aphids. *Ann. App. Biol.* **103**, 455-464. https://doi.org/10.1111/j.1744-7348.1983.tb02783.x

Watson M, Serjeant EP (1964) The effect of motley dwarf virus on yield of carrots and its transmission in the field by *Cavariella aegopodiae*. Scop. *Ann. Appl. Biol*. **53**, 77-93. https://doi.org/10.1111/j.1744-7348.1964.tb03782.x

Watson MA, Serjeant EP, Lennon EA (1964) Carrot motley dwarf and parsnip mottle viruses. *Ann. Appl. Biol.* **54**, 163-166. https://doi.org/10.1111/j.1744-7348.1964.tb01179.x

Watson MT, Falk BW (1994) Ecological and epidemiological factors affecting carrot motley dwarf development in carrots grown in the Salina Valley of California. *Plant Dis.* **78**, 477-781. https://doi.org/10.1094/PD-78-0477

Watson MT, Tian T, Estabrook E, Falk BW (1998) A small RNA resembling the beet western yellows luteovirus ST9-associated RNA is a component of the California carrot motley dwarf complex. *Phytopathology*, **88**, 164-170. https://doi.org/10.1094/PHYTO.1998.88.2.164

Whitfield AE, Falk BW, Rotenberg D (2015) Insect vector-mediated transmission of plant viruses. *Virology* **479-480**, 278-289. https://doi.org/10.1016/j.virol.2015.03.026

Wijnheijmer EHM, Brandenburg WA, Ter Borg SJ (1989) Interactions between wild and cultivated carrots (*Daucus carota* L.) in the Netherlands. *Euphytica* **40**, 147-154. https://doi.org/10.1007/BF00023309

Willenborg J, Menzel W, Vetten HJ, Maiss E (2009) Molecular characterization of two alphacryptovirus dsRNAs isolated from *Daucus carota*. *Arch. Virol.* **154,** 541-543. https://doi.org/10.1007/s00705-009-0314-z

Williams IS, Dixon AFG (2007) Life cycles and polymorphism. In: Aphids as Crop Pests, van Emden HF, Harrington R (Eds), CABI Publishing, Wallingford (UK), pp 69-81.

Wren JD, Roossinck MJ, Nelson RS, Scheets K, Palmer MW, Melcher U (2006) Plant virus biodiversity and ecology. *PLoS Biol.* **4**, e80. https://doi.org/10.1371/journal.pbio.0040080

Wu H, Fu P, Fu Q, Yu F, Peng Y (2022) Plant Virus Database: a resource for exploring the diversity of plant viruses and their interactions with hosts. *bioRxiv*. https://doi.org/10.1101/2022.03.20.485054

Xu P, Chen F, Mannas JP, Feldman T, Sumner LW, Roossinck MJ (2008) Virus infection improves drought tolerance. *New Phytol.* **180,** 911-921. https://doi.org/10.1111/j.1469-8137.2008.02627.x

Yoshida N (2020) Biological and genetic characterisation of carrot red leaf virus and its associated virus/RNA isolated from carrots in Hokkaido, Japan. *Plant Pathol.* **69**, 1379-1389. https://doi.org/10.1111/ppa.13202

Zhou CJ, Zhang XY, Liu SY, Wang Y, Li DW, Yu JL, Han CG (2017) Synergistic infection of BrYV and PEMV 2 increases the accumulations of both BrYV and BrYV-derived siRNAs in *Nicotiana benthamiana*. *Sci. Rep.* **7**, 45132. https://doi.org/10.1038/srep45132

ANNEX

Identification of two novel putative satellite RNAs with hammerhead structures in the virome of French and Spanish carrot samples

Archives of Virology (2022) 167:2287–2292
https://doi.org/10.1007/s00705-022-05538-z

BRIEF REPORT

Identification of two novel putative satellite RNAs with hammerhead structures in the virome of French and Spanish carrot samples

B. M. Babalola[1] · D. Schönegger[2] · C. Faure[2] · A. Marais[2] · A. Fraile[1] · F. Garcia-Arenal[1] · T. Candresse[2]

Received: 26 April 2022 / Accepted: 9 June 2022 / Published online: 20 July 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2022

Abstract

Carrot virome analysis using high-throughput sequencing revealed the presence of two RNA molecules with properties of satellite RNAs that are homologous to the satellite RNA of cereal yellow dwarf virus-RPV (CYDV-RPV). Satellite 1 is 298 nt long, while satellite 2 is 368 nt long. Their positive and negative genome strands contain hammerhead ribozymes similar to those found in other self-cleaving satellite RNAs. While both satellites were detected in Spanish carrot populations, only satellite 2 was found in French carrot populations. The most likely helper virus for these two satellites is carrot red leaf virus (CtRLV), which, like CYDV-RPV, is a polerovirus.

Satellite RNAs (satRNAs) are subviral agents that are dependent on a helper virus (HV) for critical functions such as replication, encapsidation, or transmission [1–3]. Unlike defective RNAs, satRNAs have no sequence similarity to their HV. Some of them are capable of modifying, positively or negatively, the disease symptoms induced in host plants by the HV, while others do not seem to substantially affect their HV or the symptoms it causes [1–3]. For example, the satRNA of tobacco ringspot virus (TRSV satRNA) reduces the symptoms of TRSV, while that of arabis mosaic virus (ArMV satRNA) intensifies ArMV symptoms [1].

Several groups of satRNAs have been distinguished based on genome size and properties [1, 2]. While some satRNAs have large genomes (in the range 1 kb and above) that contain open reading frames (ORFs), smaller non-coding linear satellite RNAs (sl-SatRNA) with genomes less than 700 nt in size have been reported with HVs of the families *Tombusviridae* and *Bromoviridae*. Another class is represented by small circular non-coding satellite RNAs (sc-SatRNA) with genomes of about 300-350 nt, which are mostly associated with helper viruses of the families *Secoviridae* (e.g., arabis mosaic virus small satellite RNA, ArMV satRNA) and *Solemoviridae*, such as the satellites associated with several members of the genus *Sobemovirus* (e.g., rice yellow mottle virus satellite, RYMV-SatRNA) [4]. The sc-SatRNAs are known to replicate by a rolling-circle mechanism and to generate unit-length genomes through self-cleavage [5, 6]. They show a high degree of secondary structure [1] and have in their positive genome strand conserved sequences that correspond to self-cleaving "hammerhead" ribozymes [6, 7]. Depending on the satellite, their negative genome strand either does not encode a ribozyme or encodes a hairpin or a hammerhead ribozyme [8]. So far, the only small satellite with hammerhead ribozymes on both genome strands and with a HV belonging to the genus *Polerovirus* is cereal yellow dwarf virus-RPV satellite RNA (CYDV-RPV satRNA) [8, 9]. While initially described as a satellite of the luteovirus barley yellow dwarf virus-RPV (BYDV-RPV), a later reclassification and splitting of the genus *Luteovirus* has led to a renaming of the HV as cereal yellow dwarf virus-RPV (CYDV-RPV) and assignment of this virus to the genus *Polerovirus* [10]. The replication of CYDV-RPV satRNA and the processing of the multimeric forms to monomeric copies by the hammerhead ribozymes encoded on its plus and minus genome strands have been studied extensively [9, 11–14]. These studies have, in particular, demonstrated that the replication of the satellite is dependent on the HV [11] and identified key

Handling Editor: Massimo Turina.

✉ T. Candresse
thierry.candresse@inrae.fr

[1] Centro de Biotecnología y Genómica de Plantas UPM-INIA, E.T.S.I. Agronómica, Alimentaria y de Biosistemas, Campus de Montegancedo, Universidad Politécnica de Madrid, Pozuelo de Alarcón, 28223 Madrid, Spain

[2] Univ. Bordeaux, INRAE, UMR BFP, CS 20032, 33882 Villenave d'Ornon Cedex, France

residues in the plus- and minus-strand ribozymes [13] as well as tentative replicase recognition sites on both satellite strands and a putative encapsidation signal [14]. Thus, a replication model has been proposed in which alternative conformations of the molecule either favor ribozyme activity or, on the contrary, favor efficient CYDV-RPV satRNA RNA replication through inhibition of ribozyme activity [12–14]. CYDV-RPV satRNA has also been shown to reduce accumulation of its HV and attenuate its symptoms [15] and to be able to be assisted by beet western yellows virus (another polerovirus) in tobacco protoplasts as well as in *Capsella bursa-pastoris* plants [16].

The present work is focused on the characterization of two new putative hammerhead-containing sc-SatRNAs (Hhsats) discovered in the virome of carrot red leaf virus-infected carrots (*Daucus carota* L.) in Spain and France.

In June 2021, five cultivated carrot (*Daucus carota* subsp. *sativa*) fields and one wild carrot (*Daucus carota* subsp. *carota*) population were sampled near Segovia (central Spain). From each population, fifty plants were sampled (irrespective of whether they showed symptoms of viral infection such as leaf reddening or yellowing) and combined in a pool. Double-stranded RNA was extracted from each pool [17], converted to complementary DNA, and sequenced (2 × 125-nt paired reads, Illumina Hiseq2500). After cleanup of the reads, contigs were assembled *de novo* using CLC Genomics Workbench v22.0 and annotated by BLASTn and BLASTx analysis against the virus section of the GenBank RefSeq database. Two groups of contigs of 293 to 368 nt with a low level of sequence similarity to CYDV-RPV satRNA (M63666) were identified in the virome of several pools. Extension of these contigs by rounds of mapping of residual reads allowed us to assemble sequences of partial multimers of two different putative small satellite RNA molecules from which unit-length monomeric sequences were derived. To confirm the presence of the two satellites in sampled carrots, a two-step RT-PCR assay was performed using specific detection primers designed using the genome sequence of each of the satellites: Sat1_F (5'-ACAGAAAACCACCCGAGTAA-3') and Sat1_R (5'-TAACCACATGGGAGTCATCCT-3'), and Sat2_F (5'-CCACCACACTCGTTTTGTG-3' and Sat2_R (5'-TCCACTTCTTCCTCGATTGAG-3'). Both primer pairs generate a 258-nt amplicon. Sat1 was present in all six sampled carrot populations, whereas Sat2 was found in only four cultivated carrot populations. Sanger sequencing of the amplicons for the two satellites showed them to be identical to the respective sequences obtained by assembly of Illumina reads. In addition, the presence of circular forms of these two molecules was verified by RT-PCR using pairs of specific divergent primers: Sat1-R1 (5′-GTCTCCTCACTTCAAAGAGTG-3′) and Sat1-F1 (5′-GCTTTACGTGTCTGTCATCAA-3′), and Sat2-R1 (5′-TACCTCGACTGATGAGTTCAA-3′) and Sat2-F1 (5′-GCACCTCGAGACACCTTTCCT-3′).

The smaller of the two molecules, which is referred to as Sat1, is 298 nt long, while the longer one, referred to as Sat2, is 368 nt long. These sequences have been deposited in the GenBank database under the accession numbers OM962993 and OM962994, respectively. The two molecules are 70.7% identical and share 84.6% (Sat1) and 69.4% (Sat2) identity with CYDV-RPV satRNA. In particular, they share two regions of high sequence similarity corresponding to the conserved hammerhead ribozymes on the plus and minus strands of their genome (corresponding to positions 1-77 and 206-260 of the Sat2 molecule). In these regions, Sat1 and Sat2 are 96% and 92.3% identical and share 93.3-96% and 87.3-90.9% identity, respectively, with the corresponding regions of CYDV-RPV satRNA (Fig. 1A and B). Given the high level of sequence similarity, the predicted folding for the plus- and minus-strand ribozymes is similar to that determined by Miller et al. [9, 12]. In both hammerheads, nucleotide differences in hairpin regions with respect to the hammerhead ribozymes of CYDV-RPV satRNA are systematically accompanied by compensatory changes on the other hairpin strand – bases 196 and 203, 194 and 229, and 195 and 228 for the minus-strand ribozyme (Fig. 1A) and bases 1 and 16, 2 and 15, and 78 and 296 for the plus strand ribozyme (Fig. 1B) – providing strong evolutionary support for the secondary structure proposed for CYDV-RPV satRNA. All other nucleotide changes as compared to CYDV-RPV satRNA hammerheads occur in loop regions (position 222 for the minus-strand ribozyme, position 21 for the plus-strand ribozyme), with the exception of several mutations (positions 287-289 and 291-290 plus 84-85) that shorten lateral stem 3 (naming according to reference [12]). In the same fashion, the double hammerhead structure that has been proposed [13] to be the active form for the processing of the plus-strand multimers shows compensatory mutations so that the potential to form stem 1 is fully conserved (not shown).

Remarkably, the region of CYDV-RPV satRNA that is essential for replication and has been proposed to be a replicase binding site or an origin of replication (positions 245-310, [14], Fig. 2A) is the most divergent region among the different satellites, showing both large length variation (66, 50, and 107 nt for CYDV-RPV satRNA, Sat1, and Sat2, respectively) and low sequence similarity (only 34% identity between Sat1 and Sat2 and 68.1% and 43.9% identity between CYDV-RPV satRNA and Sat1 and Sat2, respectively) (Fig. 2). Similarly, the CYDV-RPV satRNA 152-194 region, which has been suggested to be involved in encapsidation [14] (Fig. 2A) is poorly conserved in the two carrot satellites, again showing both length variation (43, 35, and 31 nt for CYDV-RPV satRNA, Sat1, and Sat2, respectively)

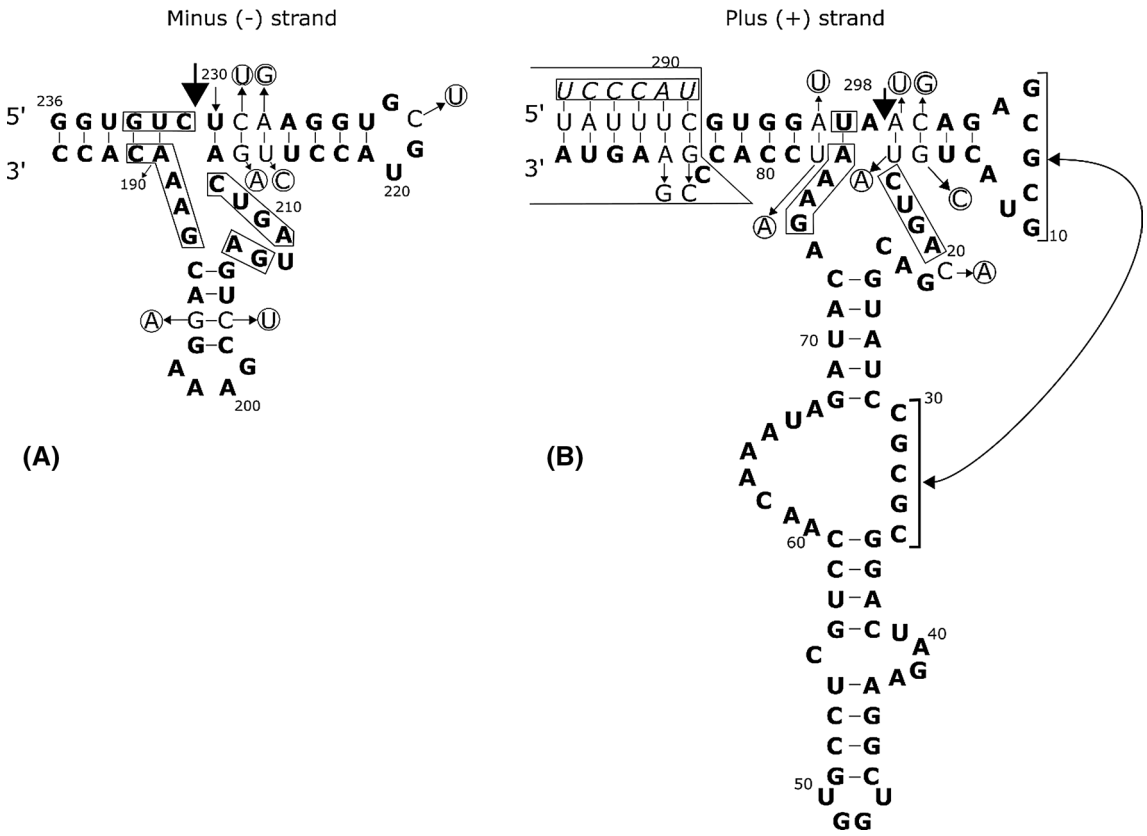


Fig. 1 Proposed secondary structures for the minus- and plus-strand hammerhead ribozymes of carrot red leaf virus satellite 1 (CtRLV Sat1) based on the structure proposed for the ribozymes of CYDV-RPV satellite RNA. Nucleotide numbering is based on the full-length Sat1 RNA (298 nt). Boxed bases are conserved among satRNA hammerhead ribozymes. Large bold arrows indicate the cleavage sites. Nucleotides in bold are conserved when compared to CYDV-RPV satRNA, while small arrows and circled letters indicate mutations as compared to CYDV-RPV satRNA. (A) Minus-strand ribozyme. Bases are numbered according to the complementary plus strand. (B) Plus-strand ribozyme. Square brackets joined by a long double arrow indicate a pseudoknot structure that is also found in CYDV-RPV satRNA

and sequence divergence (only 54.8% identity between Sat1 and Sat2 and 68.6% and 58.1% between CYDV-RPV satRNA and Sat1 and Sat2, respectively) (Fig. 2). Taken together, the secondary structure analysis of CtRLV Sat1 and Sat2 revealed three kinds of functional elements present in CYDV-RPV satRNA: (i) self-functioning ribozymes, (ii) elements that control conformational change (RNA switches, e.g., L1 and L2a, which facilitate replication), and (iii) *cis*-acting elements that interact with helper virus and/or host components, e.g., origins of replication and assembly [14].

As judged by the number of HTS reads corresponding to each molecule, the smaller Sat1 consistently showed much higher accumulation levels than the larger Sat2. For example, in the third cultivated carrot population, Sat1 represented 170,232 reads (70,155x coverage, 0.8% of total reads for the sample), while Sat2 represented only 2,218 reads (739x coverage, 0.01% of total reads), which corresponds to a 95-fold higher representation for Sat1. Given that the HTS data were generated from pools of plants, it is not known whether the two Hhsats were present together in the same plants and therefore whether this observed variation in representation reflects a true difference in accumulation in infected plants, a difference in prevalence of the two molecules in the sampled carrot populations, or a combination of both factors.

Full genome sequences could be assembled from HTS sequencing reads for five carrot populations for Sat1 but only for two populations for Sat2 due to its lower representation. The Sat1 genome sequences obtained for the five Spanish populations were completely identical. For Sat2, a single indel polymorphism was observed between the two populations, $_{123}$UUUU$_{126}$ > UUU. Analysis of HTS datasets for carrot populations collected in France in 2020 and processed in a similar fashion as the Spanish samples indicated the presence of Sat2, but not Sat1, in two populations collected in Aquitaine about 20 km apart. One of these French populations was from cultivated carrots, while the second was from wild carrots. Given the 81x-193x average coverage of Sat2 in the HTS reads for these two populations, full-length genomic sequences

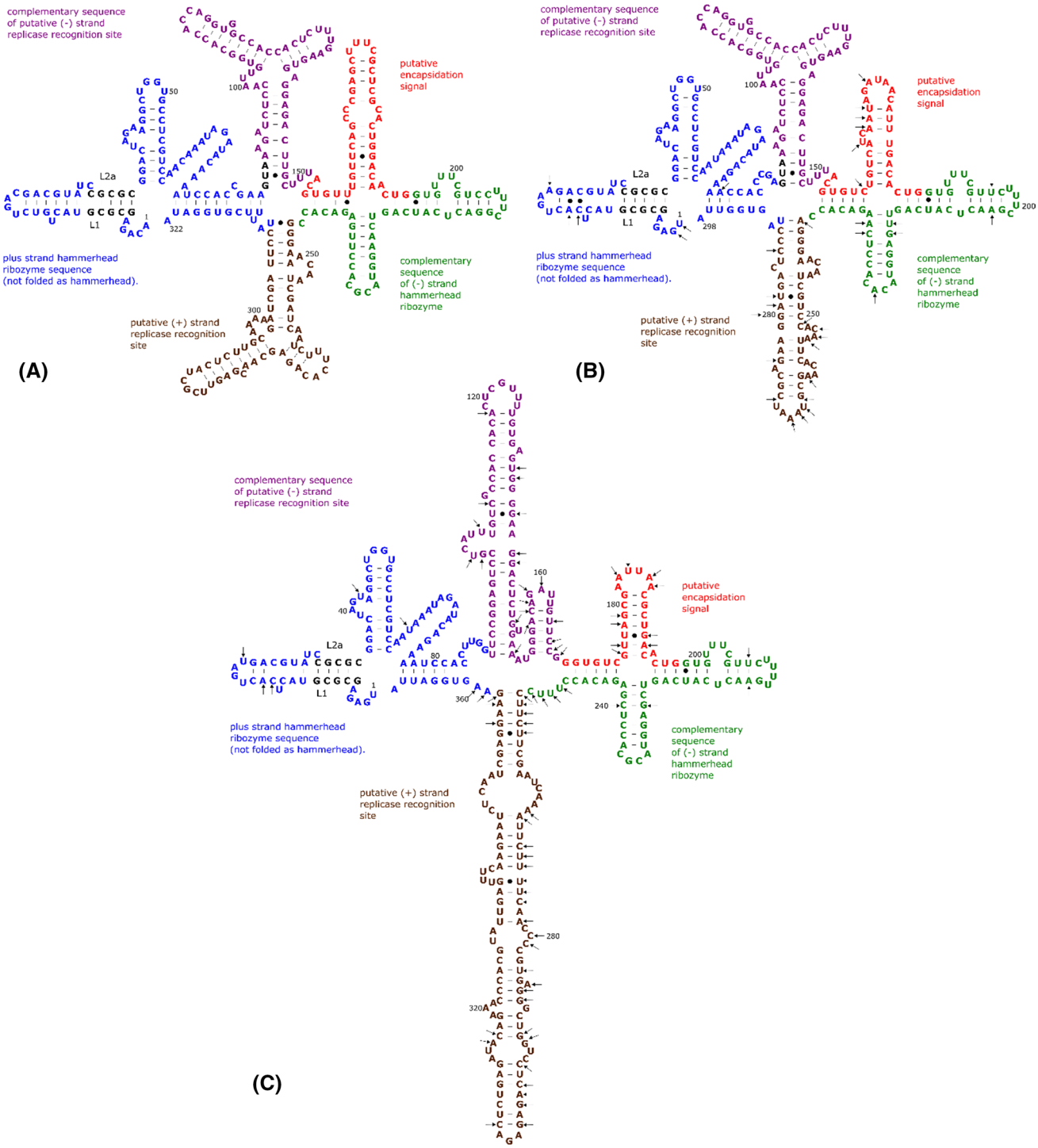


Fig. 2 Known and putative functional domains in the most stable secondary structure of the plus strand of CYDV-RPV satRNA (A), adapted from Song and Miller, 2004, CtRLV Sat1 (B) and CtRLV Sat2 (C). Structural domains are color coded for each known or proposed function as indicated, with the L1-L2a bases shown in black. Arrows indicate nucleotide differences in Sat1 and Sat2 with respect to CYDV-RPV satRNA

could be reconstructed (GenBank OM962996). The two sequences are identical and show a single polymorphism when compared to the Spanish sequences, which affects the region that is also polymorphic between the two Spanish isolates ($_{123}$U**A**UU$_{126}$). The variability between the carrot satellites found in France and Spain is therefore

low, while there is no evidence so far for the presence of Sat1 in France, despite the fact that 45 carrot populations representing a total of 2250 plants have been sampled to date. Given the differential representation of Sat1 in Spanish samples, it seems unlikely that, if present in French samples, it could have been missed during these virome studies.

HhSats have been found associated with nepoviruses in the family *Secoviridae* and sobemoviruses and polerovirus in the family *Solemoviridae*. Although HhSats have no sequence similarity to their helpers, BLASTn analysis of Sat1 and Sat2 showed significant similarity to CYDV-RPV satRNA, which has a polerovirus HV and can be assisted by at least another polerovirus [16], but detected no similarity to other HhSats. A detailed analysis of the virome of the carrot samples in which Sat1 and Sat2 were identified shows that the most likely HV candidate is carrot red leaf virus (CtRLV), which, like CYDV-RPV, is a polerovirus. In fact, CtRLV was the only member of the family *Solemoviridae* identified in these plant pools. The other viruses identified in the virome belong to genera whose members are not known to be associated with HhSats, and a potential role of carrot torradovirus 1 (CaTV1), a torradovirus of the family *Secoviridae* can be excluded, as it was not identified in the virome of the French carrot samples in which Sat2 was identified. Consequently, CtRLV appears to be the most likely HV for these two novel putative satellites, and the names CtRLV satellite 1 and CtRLV satellite 2 are therefore proposed for the two molecules identified here. To our knowledge, these results represent the first identification of satellite RNAs associated with a polerovirus under field conditions, since CYDV-RPV satRNA was first described in a laboratory sample of its HV and does not appear to have been observed again under field conditions. Since there were no specific efforts to select for symptomatic or asymptomatic plants to prepare the pools, it is not possible to know whether Sat1 or Sat2 could affect the accumulation of CtRLV or its symptomatology, in particular when considering that it was systematically found associated with several umbraviruses that are known to be synergistic with CtRLV and with other satellites or associated RNAs.

Acknowledgements The authors would like to thank the Platform GetPlaGe (INRAE, Toulouse, France) for the Illumina sequencing of the French samples.

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" (GA 813542).

Data availability The genome sequences for the hammerhead satellites reported here have been deposited in the GenBank database. The raw sequence datasets are available on request from the authors.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

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

References

1. Roossinck MJ, Sleat D, Palukaitis P (1992) Satellite RNAs of plant viruses: structures and biological effects. Microbiol Rev 56:265–279. https://doi.org/10.1128/mr.56.2.265-279.1992
2. Hu CC, Hsu YH, Lin NS (2009) Satellite RNAs and satellite viruses of plants. Viruses 1:1325–1350. https://doi.org/10.3390/v1031325
3. Gnanasekaran P, Chakraborty S (2018) Biology of viral satellites and their role in pathogenesis. Curr Opin Virol 33:96–105. https://doi.org/10.1016/j.coviro.2018.08.002
4. Briddon RW, Ghabrial S, Lin NS, Palukaitis P, Scholthof KBG, Vetten HJ (2012) Satellite and other virus-dependent nucleic acids. In: King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (eds) Virus Taxonomy: classification and nomenclature of viruses. Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier, Amsterdam, pp 1211–1219
5. Branch AD, Robertson HD (1984) A replication cycle for viroids and other small infectious RNAs. Science 223:450–455. https://doi.org/10.1126/science.6197756
6. Bruening G, Passmore BK, van Tol H, Buzayan JM, Feldstein PA (1991) Replication of plant virus satellite RNA: evidence favors transcription of circular templates of both polarities. Mol Plant Microb Interact 4:219–225. https://doi.org/10.1094/mpmi-4-219
7. Bruening G (1989) Compilation of self-cleaving sequences from plant virus satellite RNAs and other sources. Methods Enzymol 180:546–558. https://doi.org/10.1016/0076-6879(89)80123-5
8. Navarro B, Rubino L, di Serio F (2017) Small Circular Satellite RNAs. In: Hadidi A, Flores R, Randles JW, Palukaitis P (eds) Viroids and satellites. Academic Press, pp 659–669. https://doi.org/10.1016/B978-0-12-801498-1.00061-9
9. Miller WA, Hercus T, Waterhouse PM, Gerlach WL (1991) A satellite RNA of barley yellow dwarf virus contains a novel hammerhead structure in the self-cleavage domain. Virology 183:711–720. https://doi.org/10.1016/0042-6822(91)91000-7
10. Domier LL (2012) Luteoviridae. In: King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (eds) Virus Taxonomy: classification and nomenclature of viruses. Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier, Amsterdam, pp 1045–1053
11. Silver SL, Rasochova L, Dinesh-Kumar SP, Miller WA (1994) Replication of barley yellow dwarf virus satellite RNA transcripts in oat protoplasts. Virology 198:331–335. https://doi.org/10.1006/viro.1994.1036
12. Miller WA, Silver SL (1991) Alternative tertiary structure attenuates self-cleavage of the ribozyme in the satellite RNA of barley yellow dwarf virus. Nucleic Acids Res 19:5313–5320. https://doi.org/10.1093/nar/19.19.5313
13. Song SI, Silver SL, Aulik MA, Rasochova L, Mohan BR, Miller WA (1999) Satellite cereal yellow dwarf virus-RPV (satRPV) RNA requires a double hammerhead for self-cleavage and an alternative structure for replication. J Mol Biol 293:781–793. https://doi.org/10.1006/jmbi.1999.3169

14. Song SI, Miller WA (2004) *Cis* and *trans* requirements for rolling circle replication of a satellite RNA. J Virol 78:3072–3082. https://doi.org/10.1128/jvi.78.6.3072-3082.2004
15. Rasochová L, Miller WA (1996) Satellite RNA of barley yellow dwarf-RPV virus reduces accumulation of RPV helper virus RNA and attenuates RPV symptoms in oats. Mol Plant Microb Interact 9:646–650. https://doi.org/10.1094/MPMI-9-0646
16. Rasochová L, Passmore BK, Falk BW, Miller WA (1997) The satellite RNA of barley yellow dwarf virus-RPV is supported by beet western yellows virus in dicotyledonous protoplasts and plants. Virology 231:182–191. https://doi.org/10.1006/viro.1997.8532
17. Marais A, Faure C, Bergey B, Candresse T (2018) Viral double-stranded RNAs (dsRNAs) from plants: alternative nucleic acid substrates for high-throughput sequencing. Methods Mol Biol 1746:45–53. https://doi.org/10.1007/978-1-4939-7683-6_4

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