Phytopathology®
REVIEW

Virion-Associated Nucleic Acid-Based Metagenomics: A Decade of Advances in Molecular Characterization of Plant Viruses

Oumaima Moubset,[1,2] Sarah François,[3] François Maclot,[4] Essowè Palanga,[5] Charlotte Julian,[1,2] Lisa Claude,[1,2] Emmanuel Fernandez,[1,2] Philippe Rott,[1,2] Jean-Heinrich Daugrois,[1,2] Aymeric Antoine-Lorquin,[6] Pauline Bernardo,[7] Arnaud G. Blouin,[4,8] Coline Temple,[4] Simona Kraberger,[9] Rafaela S. Fontenele,[9] Gordon W. Harkins,[10] Yuxin Ma,[11] Armelle Marais,[11] Thierry Candresse,[11] Sélim Ben Chéhida,[12] Pierre Lefeuvre,[12] Jean-Michel Lett,[12] Arvind Varsani,[9,13] Sébastien Massart,[4] Mylène Ogliastro,[3] Darren P. Martin,[14] Denis Filloux,[1,2] and Philippe Roumagnac[1,2,†]

[1] CIRAD, UMR PHIM, 34090 Montpellier, France
[2] PHIM Plant Health Institute, Univ Montpellier, CIRAD, INRAE, Institut Agro, IRD, Montpellier, France
[3] Department of Zoology, University of Oxford, U.K.
[4] Plant Pathology Laboratory, Terra, Gembloux Agro-Bio Tech, Liège University, Gembloux, Belgium
[5] Institut Togolais de Recherche Agronomique (ITRA-CRASS), B.P. 129, Kara, Togo
[6] INRA-Université de Montpellier UMR DGIMI 34095 Montpellier, France
[7] Enza Zaden, Haling 1-E, 1602 DB Enkhuizen, The Netherlands
[8] Plant Protection Department, Agroscope, 1260, Nyon, Switzerland
[9] The Biodesign Center for Fundamental and Applied Microbiomics, Center for Evolution and Medicine, School of Life Sciences, Arizona State University, Tempe, AZ, U.S.A.
[10] South African Medical Research Council Capacity Development Unit, South African National Bioinformatics, Institute, University of the Western Cape, South Africa
[11] Univ. Bordeaux, INRAE, UMR BFP, 33140 Villenave d'Ornon, France
[12] CIRAD, UMR PVBMT, F-97410 St Pierre, La Réunion, France
[13] Structural Biology Research Unit, Department of Integrative Biomedical Sciences, University of Cape Town, Observatory, Cape Town, South Africa
[14] Division of Computational Biology, Department of Integrative Biomedical Sciences, Institute of Infectious Diseases and Molecular Medicine, University of Cape Town, Cape Town, South Africa

Accepted for publication 16 June 2022.

Abstract

Over the last decade, viral metagenomic studies have resulted in the discovery of thousands of previously unknown viruses. These studies are likely to play a pivotal role in obtaining an accurate and robust understanding of how viruses affect the stability and productivity of ecosystems. Among the metagenomics-based approaches that have been developed since the beginning of the 21st century, shotgun metagenomics applied specifically to virion-associated nucleic acids (VANA) has been used to disentangle the diversity of the viral world. We summarize herein the results of 24 VANA-based studies, focusing on plant and insect samples conducted over the last decade (2010 to 2020). Collectively, viruses from 85 different families were reliably detected in these studies, including capsidless RNA viruses that replicate in fungi, oomycetes, and plants. Finally, strengths and weaknesses of the VANA approach are summarized and perspectives of applications in detection, epidemiological surveillance, environmental monitoring, and ecology of plant viruses are provided.

Keywords: ecology, pathogen detection, virology

†Corresponding author: P. Roumagnac; philippe.roumagnac@cirad.fr

Funding: O. Moubset and S. Ben Chéhida are recipient of Ph.D. fellowships from the Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD) and Agence Nationale de la Recherche (ANR) (Phytovirus project ANR-19-CE35-0008-02). P. Lefeuvre and J.-M. Lett were supported by the European Regional Development Fund (contract GURDT I2016-1731-0006632), the Conseil Régional de La Réunion, and CIRAD. A. Blouin and S. Massart were supported by the project SEVIPLANT from Federal Public Service, Public Health, Belgium (grant agreement RT 18/3). C. Temple was supported by INEXTVIR Project from European Union's Horizon Europe Marie Sklodowska-Curie (grant agreement 813542).

e-Xtra: Supplementary material is available online.

The author(s) declare no conflict of interest.

In 2005, Patrick D. Schloss and Jo Handelsman metaphorically compared the problem of bacterial "unculturability" to the Gordian knot of Greek legend that Alexander the Great, the king of the ancient Greek kingdom of Macedon, unfastened with his sword (Schloss and Handelsman 2005). They proposed that metagenomic studies using the 16S ribosomal RNA (16S rRNA) gene sequences to disentangle bacterial diversity could effectively slice through this metaphorical knot (Schloss and Handelsman 2005). Since then, the extensive use of the 16S rRNA gene sequences has achieved just that. However, another Gordian knot was left largely untouched: that of viral diversity for which an analogue of a universally conserved prokaryotic 16S rRNA gene does not exist.

To overcome this complex problem, virologists needed to develop metagenomics-based approaches that target not only diverse viral genes but also different classes of nucleic acids (Greninger 2018; Maclot et al. 2020; Roossinck 2012; Roossinck et al. 2015; Roux et al. 2019), including total RNA or DNA; the encapsidated fractions of DNA and/or RNA; double-stranded cellular RNAs (dsRNA); circular DNA, and virus-derived small interfering RNAs (siRNAs). These classes of viral nucleic acids are located in various compartments of the plant cell (Fig. 1). One of these approaches, so-called "viromics", based on shotgun metagenomics specifically applied to virion-associated nucleic acids (VANA) has gained popularity over the past decade and paved the way toward disentangling the still formidable knot of global viral diversity (Roux et al. 2019; Sommers et al. 2021).

Viromics and other viral metagenomics approaches, like total RNA-based metagenomics and metatranscriptomics (Neri et al. 2022; Shi et al. 2016; Sommers et al. 2021; Zhang et al. 2018), revealed remarkable levels of RNA and DNA virus diversity in various animal, plant, and environmental samples (Call et al. 2021; Camarillo-Guerrero et al. 2021; Dávila-Ramos et al. 2019; Dutilh et al. 2021; Hasiów-Jaroszewska et al. 2021; Maclot et al. 2020; Obbard 2018). This resulted in a virtual "tsunami" of complete or partial genome sequence deposits to the comprehensive NCBI GenBank database (Dutilh et al. 2021; Greninger 2018; Roux et al. 2019). Specifically, whereas in 2010 only 84 metagenomics-derived partial or complete viral genome sequences were available in the database, by 2016 this had increased to 35,000 and by 2018 the database contained over 755,000 publically available sequences (Roux et al. 2019). This exponential growth of viral sequences in the public database is also apparent within the NCBI RefSeq database,

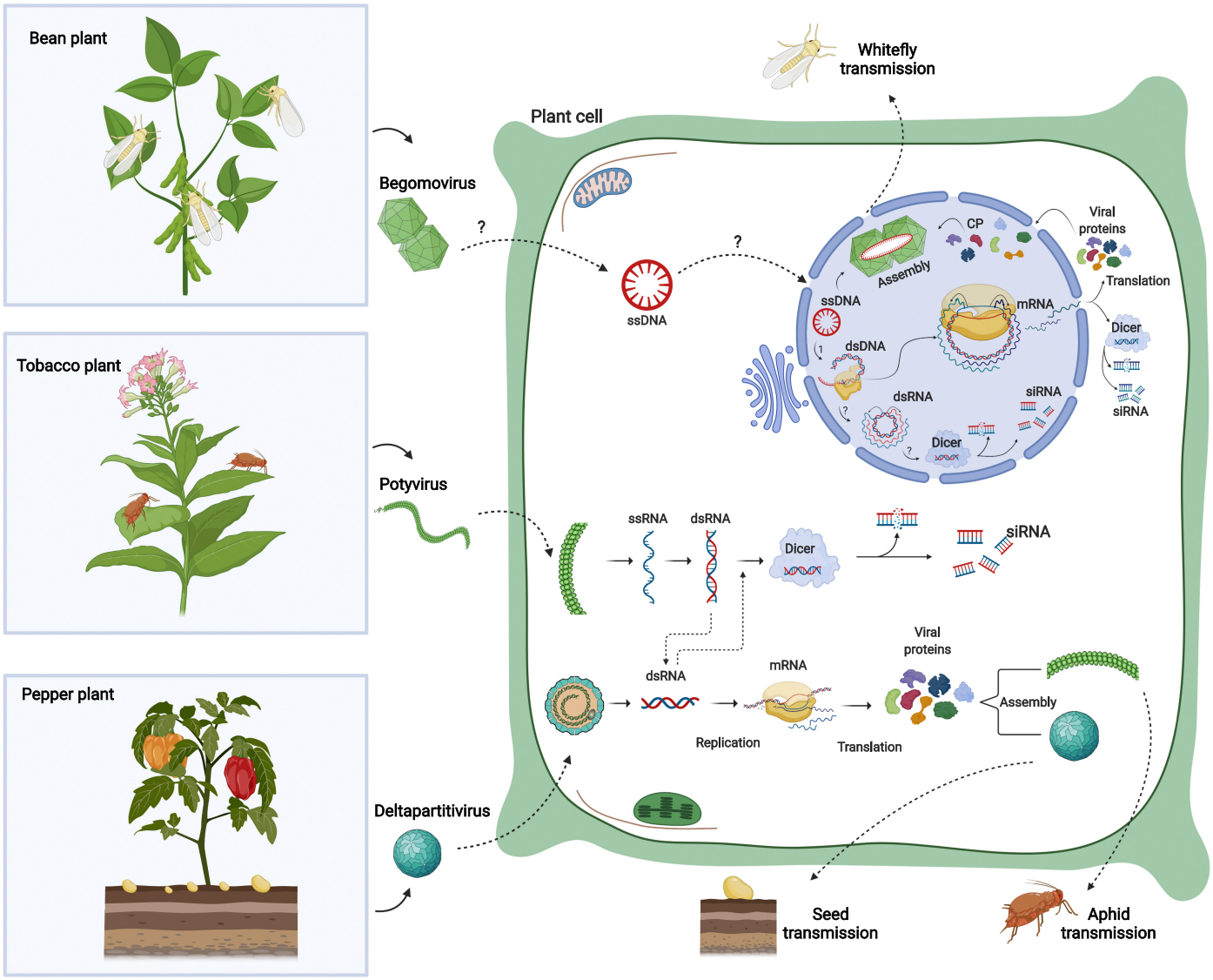


FIGURE 1
Cartoon representation of a plant cell depicting various compartments where virus-associated nucleic acids are located during virus replication and RNA silencing (sequence-dependent system for controlling gene expression).

which contains one reference genome per species. For example, between August 2020 and August 2021 *Riboviria* reference genome sequences within this database increased from 3,922, to 6,630, an increase of 70% in a single year.

Nevertheless, although the number of complete viral genomes retrieved from human, vertebrate, or environmental samples has undergone an exponential increase (Edgar et al. 2022), the complete viral genomes derived from plant samples have seen a more modest magnification (Maclot et al. 2020). Additionally, a huge proportion of available plant virus genome sequences are for viruses infecting economically important cultivated crop species displaying disease symptoms (Roossinck 2012; Roossinck et al. 2015; Stobbe and Roossinck 2014; Wren et al. 2006), which represents only a small fraction of all existing plant species.

Considering the paucity of data on plant virus diversity, the Plant Virus Biodiversity and Ecology (PVBE) project was initiated in 2005. The aim of this project was to use metagenomic approaches to better understand the biodiversity of viruses affecting vascular plants (Wren et al. 2006), with a particular focus on viruses infecting noncultivated plants. In this review, we first retrace the historical steps that led to the advent, use and application of the VANA approach to plant and insect samples. We then focus on the progress made on detection, epidemiological surveillance, environmental monitoring, and ecology of plant viruses using this approach.

VANA-Based Metagenomics

Historical foundations of the VANA approach

Within the frame of the PVBE project, the Nature Conservancy's Tallgrass Prairie Preserve (TPP) of Oklahoma (USA) was selected as a sampling pilot site because of its rich, non-cultivated flora. Plants were collected in most areas of the TPP irrespective of possible viral infection symptoms, and, for the first time in the history of plant virology, the samples were processed with a virus-like particle-viral nucleic acid extraction (VLP-VNA) metagenomic method. The VLP-VNA metagenomic method included several steps of centrifugation, ultracentrifugation and DNAse treatment of individual plant samples. These first steps were followed by the extraction of viral nucleic acids from VLP-containing pellets using a classical proteinase K phenol/ethanol extraction protocol (Melcher et al. 2008). The resulting nucleic acid molecules were amplified using a random PCR strategy and amplicons were cloned and Sanger sequenced (Melcher et al. 2008). When applied to investigate 95 individual specimens from 52 TPP plant species, this VLP-VNA approach resulted in successful detection and characterization of both known and unknown viruses belonging to several families (including *Caulimoviridae*, *Flexiviridae*, *Chrysoviridae*, *Comoviridae*, and *Tymoviridae*) (Muthukumar et al. 2009). However, most (34 to 67%) of the obtained VLP-VNA-associated sequences revealed by the PVBE project had no detectable similarity to any known virus sequences in the NCBI database and were therefore categorized as "dark matter" of indeterminate origin. Overall, viral dark matter still represents a substantial proportion of metagenomic data, particularly from undersampled hosts, which still strongly biases our view of the diversity and ecological roles of viruses (Krishnamurthy and Wang 2017).

Although the VLP-VNA metagenomic approach substantially expanded our knowledge of plant virus diversity within natural ecosystems, the method was cumbersome and its reliance on Sanger sequencing meant that it was relatively low throughput and suitable only to analyze tens and not hundreds of plant samples with approximately 100 sequences generated for each sample. However, almost concomitantly with the development of the VLP-VNA approach, the high throughput-sequencing (HTS) era was beginning, which provided an unprecedented opportunity to tremendously increase the throughput of VLP-VNA-like metagenomic approaches.

Hence, in 2009 and still within the framework of the PVBE project, Marilyn Roossinck and colleagues developed the so-called Ecogenomics approach (Roossinck et al. 2010), thus accelerating the pace of RNA plant virus discovery and characterization. This pioneering approach was based on the isolation of dsRNA, a hallmark of RNA virus infection (Dodds et al. 1984). The first key innovation of this new metagenomics approach was an astute nucleic acid amplification protocol that involved the use of a set of 96 tags. These tags allowed the low-cost sequencing of pooled samples and enabled the linking of sequences to their original geo-referenced hosts. The second key innovation was the use of the Roche 454 GS FLX pyrosequencing device to yield hundreds of thousands of reads from pooled samples in a single sequencing run.

This first Ecogenomics study revealed the presence of identifiable viral sequences in 70% of the 473 analyzed individual plant samples in two different ecosystems: the TPP and the "Área de Conservación Guanacaste" in northwestern Costa Rica (Roossinck et al. 2010). The viruses identified included members of the *Bromoviridae*, *Caulimoviridae*, *Chrysoviridae*, *Closteroviridae*, *Endornaviridae*, *Solemoviridae*, *Narnaviridae*, *Partitiviridae*, *Potyviridae*, *Totiviridae*, and *Tymoviridae* families (Roossinck et al. 2010).

While the prototype Ecogenomics study represented a major milestone in the field, its main shortcoming was that the dsRNA extraction protocol was cumbersome and was possibly biased towards the detection of positive-sense RNA viruses (Roossinck et al. 2015).

Development of the VANA approach

A VANA approach aimed at plant samples was then developed by combining the respective strengths of the VLP-VNA and Ecogenomics approaches (François et al. 2018a). For VANA metagenomics, virus particles are semipurified and a primer-tagging strategy is applied to trace back resulting viral sequences to their hosts and georeferenced origin, thus reducing the cost of individual sample sequencing (Roossinck et al. 2010). Furthermore, the use of filtration, ultracentrifugation and RNase/DNase digestion cocktails substantially reduces host-derived sequence reads (Breitbart et al. 2003; Victoria et al. 2008).

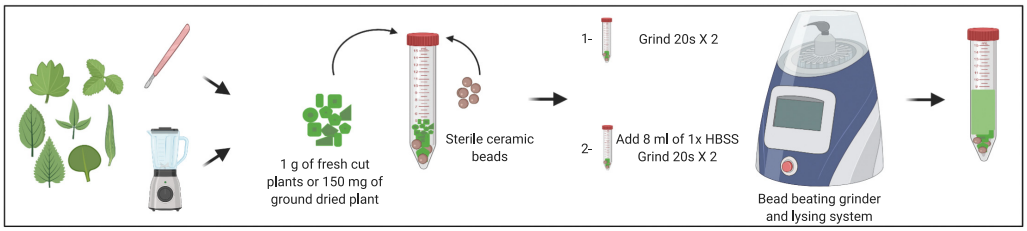
Crucially, the VANA approach is suited to the detection of both DNA and RNA viruses. It is based on five main steps that aim at removing host nucleic acids so as to maximize the yield of VANA (Fig. 2): (i) isolation of viral particles from individual or pooled plant samples using centrifugation and filtration techniques; (ii) concentration of viral particles present in the cleaned plant homogenates by ultracentrifugation; (iii) removal of contaminating nonencapsidated nucleic acids by DNase and RNase digestion treatments; (iv) extraction of encapsidated DNA and RNA molecules resistant to the DNAse and RNAse treatments, reverse transcription, Klenow fragment treatment, and amplification of the viral DNA and RNA using barcoded PCR primers (Fig. 3); and (v) sequencing of amplicons using HTS technologies and bioinformatics analysis of data (see next paragraph). The protocol of the VANA approach is comprehensively detailed in François et al. (2018a).

Development of VANA-associated bioinformatics tools

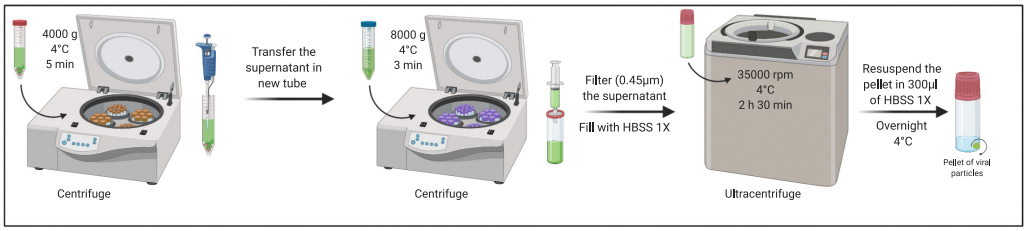
Multiple bioinformatics tools have been developed to identify sequence reads derived from viruses and further assign them to taxonomic units: either formally classified families, genera and species, or less-formal, operationally useful groupings called operational taxonomic units (OTUs) that are based on pairwise nucleotide or amino acid sequence similarity thresholds. VANA-associated bioinformatics workflows generally involve several steps (François et al. 2018a), including initially a "read-cleaning" step to verify the number of reads and evaluates their average quality using, for example, the FastQC computer program (Fig. 3). Next, a "demultiplexing"

FIGURE 2
Overall laboratory protocol of the virion-associated nucleic acids (VANA) approach that is based on six main steps that aim at removing host nucleic acids so as to maximize the yield of VANA. Step I (lysis of plant cells): individual or pooled plant samples are ground using sterile ceramic beads and tissue homogenizer. Step II (clarification of viral particles): viral particles from plant samples are isolated using centrifugation and filtration techniques. Step III (VANA extraction): contaminating nonencapsidated nucleic acids are removed by DNase and RNase digestion treatments and encapsidated DNA and RNA molecules resistant to the DNAse and RNAse treatments are further extracted. Step IV (reverse transcription): reverse transcription is carried out using homemade Dodeca linkers that enable barcoding each sample. Step V (cDNA purification and Klenow amplification): double-stranded DNA is synthetized from single-stranded DNA using large (Klenow) fragment DNA polymerase and the Dodeca linkers used during the reverse transcriptase (RT) step. Step VI (PCR amplification and verification of PCR products): double-stranded DNAs are further amplified using one multiplex identifier primer. Pools of up to 96 multiplex identifier amplicons can then be mixed and calibrated. The most recent detailed protocol, including materials and methods, is described in François et al. (2018a).

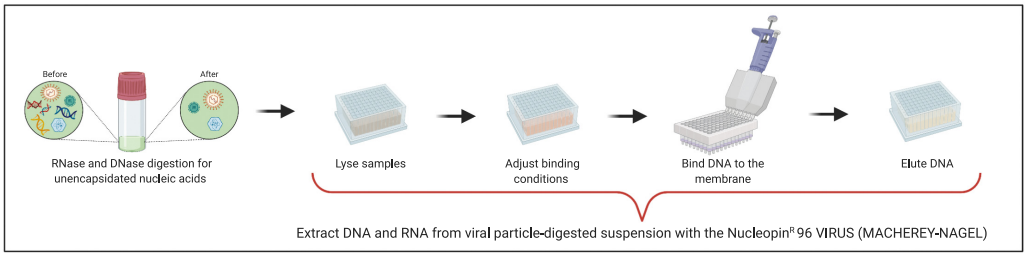
I - Lysis of plant cells



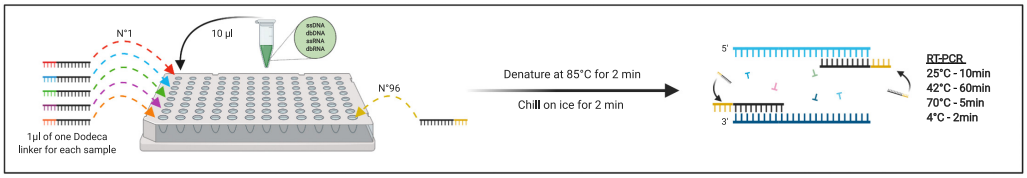
II- Clarification of viral particles



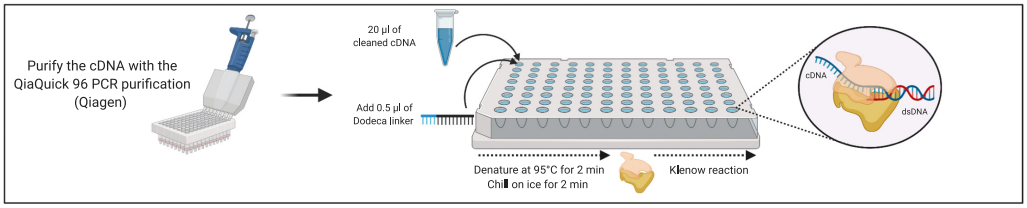
III - Viral nucleic acid extraction



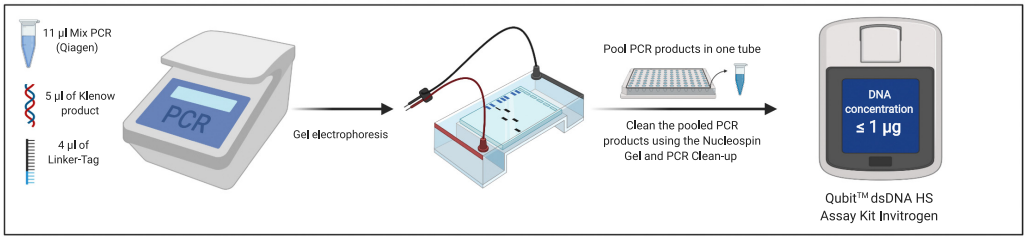
IV - Reverse transcription



V - cDNA purification and Klenow amplification



V - PCR amplification and verification of PCR products



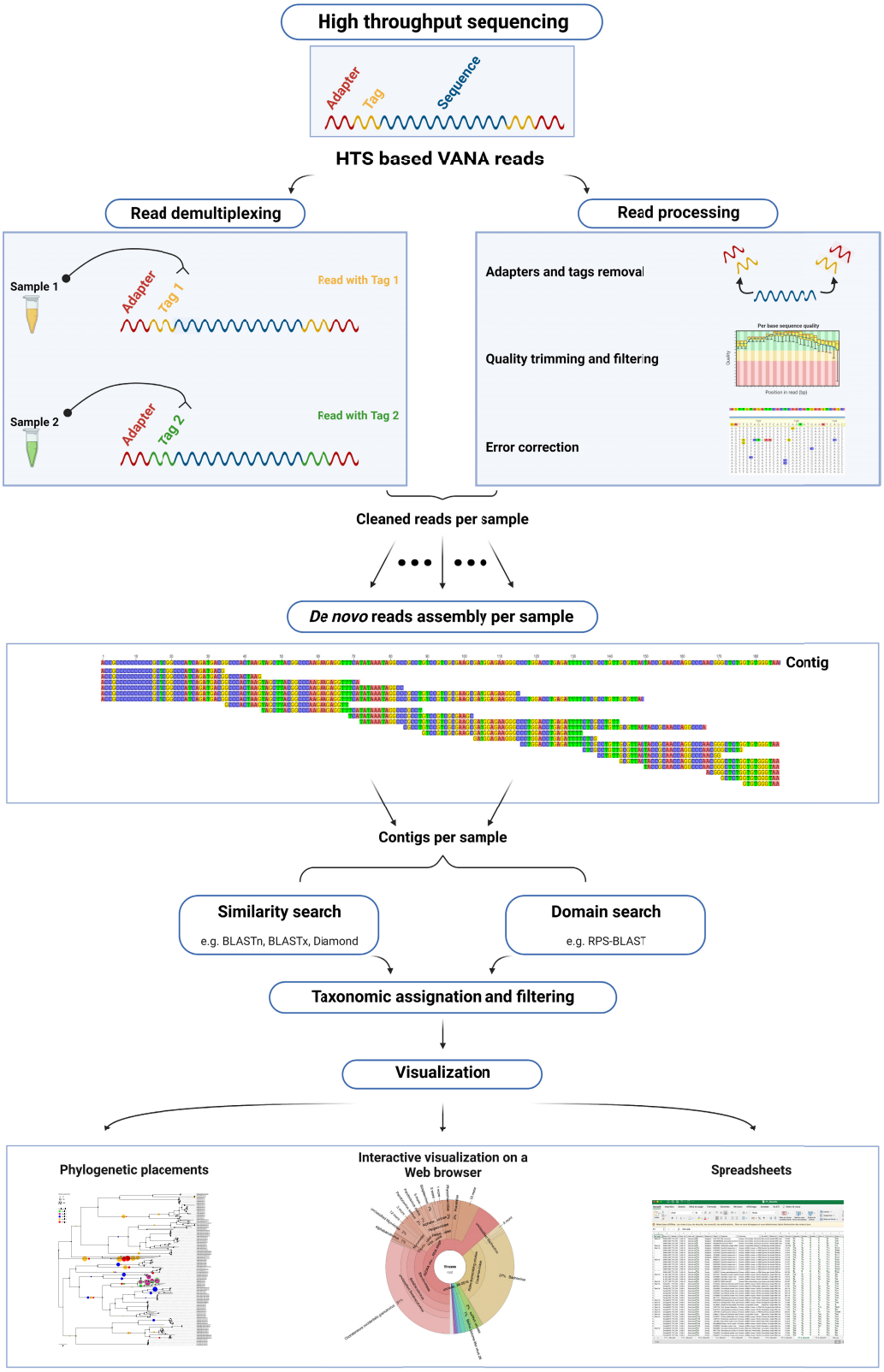


FIGURE 3
General workflow of the bioinformatics analysis of virion-associated nucleic acids (VANA) high-throughput sequencing data. The PCR step (step VI of Fig. 2) yields amplicons that are all tagged at both extremities with the same multiplex identifier primer (i.e., tag depicted in yellow in the high throughput-sequencing (HTS)-based VANA reads located at the top of the figure). HTS-based VANA reads are initially processed using two parallel steps, including on one side a "demultiplexing" step identifying each PCR primer tag in each raw read and on the other side a "read-cleaning" step (adapter and tags removal, quality trimming and filtering, and error correction). Cleaned reads are then assembled into longer continuous sequences (contigs). Finally, BLASTn, BLASTx, and/or DIAMOND and RPS-BLAST can be applied to taxonomically assign these contigs. Taxonomic assignment details can be visualized using either interactive visualization on a web browser or spreadsheets. The contigs can also be used for phylogenetic placement analyses.

step identifying each PCR primer tag in each raw read using, for example, the "agrep" Unix command (Wu and Manber 1992), to assign reads to the particular samples from which they originated. This is followed by an adaptor-removal step that discards sequencing instrument-specific adaptor sequences (such as those of the Illumina sequencers) and PCR primer sequences. Reads are then "cleaned" with a read quality filtering step, such as that implemented in Cutadapt software (Martin 2011), that removes short sequences (e.g., <15 nt) and individual sequence regions with low quality scores (for example with <q30). The cleaned reads can then be assembled into longer continuous sequences (contigs) using, for example, SPAdes or MEGAHIT assemblers (Bankevich et al. 2012). Finally, BLASTn, BLASTx (Altschul et al. 1990), and/or DIAMOND (Buchfink et al. 2021), sometimes associated with mapping tools (e.g., BOWTIE2 or BWA) (Langmead 2010), can be applied to compare contigs and remaining nonassembled reads to known viral sequences that are present in a reference database such as GenBank. This final step enables a tentative taxonomic classification of the reads and contigs by identifying those that match a virus with, for example, a specific E-value threshold. While a 10^{-3} E-value threshold has generally been widely used in VANA studies, other more stringent criteria have also been used (e.g., an E-value threshold of 10^{-4} in Ma et al. 2019). Reads assignations can then be differently visualized (Microsoft Excel or LibreOffice spreadsheets, interactive metagenomic visualization in a Web browser [Ondov et al. 2011], phylogenetic placements, etc.) (Fig. 3). A conservative approach is to classify reads and contigs to the family level to avoid "over-counting" the numbers of different taxa within a set of samples (Bernardo et al. 2018). Another possible approach focuses on alignments of conserved viral protein motifs as determined using reverse-position-specific (RPS)-BLAST against the Pfam database [e.g., RNA-dependent RNA polymerases (RdRp)], and use of phylogenetic clustering of the sequences in these alignments to define OTUs that serve as a credible proxy of viral species (Ma et al. 2019).

The 24 VANA-based studies investigated herein and conducted over the past decade yielded a large collection of reads, contigs and sample-associated metadata. The accelerating pace of data production is resulting in increasingly difficult database storage and retrieval problems. In response to these challenges, new tools such as metaXplor (Sempéré et al. 2021) have been developed to expedite searches of the sequence datasets generated by VANA studies using similarity-based search algorithms and phylogenetic tools. Along with these sequence-centric functionalities, metaXplor also facilitates the storage, retrieval, and re-analysis of data from previous VANA studies (Sempéré et al. 2021).

Detection and Identification of Plant and Insect Viruses

Viruses detected using the VANA approach

Since the VANA approach involves the removal of nonencapsidated nucleic acids by DNase and RNase digestion, viroids and other agents that are not encapsidated (or whose particles are unstable) should theoretically not be detected by this approach (Maclot et al. 2020). However, viruses in the *Closteroviridae* and *Virgaviridae* families, which are known to have quite labile, RNA-sensitive particles and members of the *Endornaviridae* family that are capsidless and produce only membranous vesicles containing the viral nucleic acids, have both been found in abundance in several independent VANA studies (Bernardo et al. 2018; Koloniuk et al. 2018; Kwibuka et al. 2021). Consequently, the VANA approach is not limited to the detection of virion-producing agents.

To further test the prediction that the VANA approach is poorly suited to the detection of viroids and viruses that do not produce capsids, we present here an inventory of all the viruses in viral families detected in 24 VANA-based metagenomics projects. These 24 projects collectively processed 9,752 plant and invertebrate samples, representing 25 cultivated plant species, 519 uncultivated plant species, 16 insect species and 2 snail species from 17 different countries (Table 1; Supplementary Table S1). The examined datasets were produced between 2010 and 2019 and represent a total of 699,548 viral contigs (>50 nt) that were classifiable at the family level using BLASTx searches (Table 1; Supplementary Table S1).

One of the critical aspects of using BLAST searches to assign contigs to viral families is the minimum length of the contig being queried. We therefore conducted a simulation experiment using 38 viral contigs (length > 1,000 nt) representing novel viruses that were assigned to 16 virus families (Supplementary Table S2). The sequences of these 38 contigs are also listed in Supplementary Table S2. In order to determine the minimum contig length required for a reliable taxonomic assignation, the 38 contigs were truncated from their first base into shorter subfragments of 50, 100, 200, 500, and 1,000 nt that were further analyzed using BLASTx searches against the nr GenBank database. The nucleotide sequences of all subfragments of 50, 100, 200, 500, and 1,000 nt are listed in Supplementary Table S2. The taxonomic assignations at the family, genus and species levels (as determined by best BlastX hit) were compared for each complete contig and for it's corresponding 50 to 1,000 nt long subfragments (Table 2).

For each length group, we thus determined the proportion of subfragments for which (i) similar viral BLASTx assignations were made to that of the full-length sequence from which they originated; (ii) different viral BLASTx assignations were made; and (iii) no BLASTx taxonomic assignations with an E-value < 0.001 were returned (Table 2).

For all query contig lengths, the accuracy of the BLASTx family and genus assignations were high, i.e., >97.0 and >86.7% for family and genus assignations, respectively (Table 2). However, the assignation accuracy at species-level (41.0 to 65.2%) was substantially lower. The proportion of unassigned contigs also increased with decreasing subfragment length (from 3.1% for 1,000 nt long subfragments to 99.8% for 50 nt long subfragments). While the proportion of unassigned subfragments was only slightly higher for the 500 nt fragments than for the 1,000 nt fragments, it increased substantially when the fragment size was decreased from 500 to 50 nt (Table 2). For all but nine of the 500 nt contigs, the BLASTx assignations were the same as for those of the 38 original sequences (Table 2). Interestingly, three of these nine misassigned contigs corresponded to artefactual chimeric carlavirus (family *Betaflexiviridae*) and potyvirus (family *Potyviridae*) reads from viruses coinfecting the analyzed plant. In addition, four out of the nine misassigned contigs were alternatively assigned to very closely related taxonomic clades, i.e., to the *Betaflexiviridae* or *Alphaflexiviridae* families for two of four contigs, and to the *Solemoviridae*/*Polerovirus*–*Tombusviridae*/*Luteovirus* genera for the remaining two contigs. Finally, two of nine of the incorrectly assigned 500 nt contigs were obtained from Novel virus_22 (Supplementary Table S2), which shared high nucleotide identity with an unclassified virus isolated from a cassava plant (cassava torrado-like virus, GenBank accession number UAW09555.1) that combines genes from viruses belonging to the families *Secoviridae* and *Solemoviridae*.

Altogether, this thorough analysis of the nine outlier contigs revealed that the BLASTx "misassignations" were due to both the presence of chimeric viruses biologically or computationally created, and/or to the imprecisely-defined phylogenetic affinities that exist between some virus families/genera (e.g., *Alphaflexiviridae*/*Betaflexiviridae*, *Solemoviridae*/*Polerovirus*–*Tombusviridae*/*Luteovirus*). We conclude that these instances of incorrect assignation are not likely to bias the global inventory of viral families detected by the VANA approach. Therefore, we propose that only contigs with lengths >500 nt should be taxonomically assigned if highly accurate family-level assignations are desired.

TABLE 1
Characteristics of the 24 virion-associated nucleic acids (VANA)-based metagenomics projects considered herein

Project name	Sequencing technology	Host (number of species)	Origin of sample (number of species)	Collection date	Country	Number of contigs (Mbases)	Number of virus contigs (Mbases)	Percentage of virus contigs
C1	Roche 454	Plant (150)	Cultivated (11), uncultivated areas (139)	2010	France	167,096 (44.56)	3,783 (1.79)	2.2%
F1	Roche 454	Plant (86)	Cultivated (6), uncultivated areas (80)	2010	South Africa	313,222 (67.37)	5,279 (1.88)	1.7%
C2	Roche 454	Plant (155)	Cultivated (12), uncultivated areas (143)	2012	France, South Africa	309,155 (66.70)	4,851 (1.85)	1.6%
F2	Roche 454	Plant (99)	Cultivated (10), uncultivated areas (89)	2012	France, South Africa	415,566 (87.84)	2,653 (0.83)	0.6%
N1	Roche 454	Plant	Cultivated areas (1)	2013	Burkina Faso	224,676 (46.12)	7,933 (2.57)	3.5%
RBR	Illumina	Plant (135)	Herbaria (94), uncultivated areas (41)	2013	France, South Africa	3,264,762 (335.74)	4,089 (0.81)	0.13%
S1	Roche 454	Plant (6)	Cultivated areas (6)	2013	France, Guadeloupe Island, Madeira Island, Reunion Island	68,559 (18.68)	4,800 (2.15)	7.0%
S2	Roche 454	Plant (6)	Cultivated areas (6)	2013	France, Guadeloupe Island, Madeira Island, Reunion Island	274,907 (71.34)	837 (0.31)	0.3%
BER	Roche 454	Plant (128)	Herbaria (77), cultivated (1), uncultivated areas (50)	2014	France, South Africa, USA	388,223 (78.04)	18,773 (5.08)	4.8%
RU	Roche 454	Plant (4)	Cultivated areas (4)	2014	Burkina Faso, Madeira Island, Reunion Island	123,318 (24.26)	2,385 (0.77)	1.9%
S3	Roche 454	Plant (6)	Cultivated areas (6)	2014	Azores, France, Guadeloupe Island, Madeira Island, Reunion Island	390,462 (81.34)	6,737 (1.91)	1.7%
S4	Roche 454	Plant (3)	Cultivated areas (3)	2014	Azores, France, Guadeloupe Island, Madeira Island	356,443 (81.97)	3,895 (1.24)	1.1%
MM	Illumina	Insect (16), snail (2), plant (7)	Cultivated (4), uncultivated areas (3)	2015	France, Madagascar, Mayotte Island, Reunion Island	1,387,606 (171.67)	11,804 (5.05)	0.9%
S6	Roche 454	Plant (6)	Cultivated areas (6)	2015	Azores, France, Guadeloupe Island, Madeira Island	98,525 (20.44)	5,633 (1.81)	5.7%
SM	Illumina	Insect (16), snail (2), plant (3)	Cultivated (1), uncultivated areas (2)	2015	France	406,798 (64.77)	9,392 (5.08)	2.3%
YY1	Illumina	Plant (2)	Cultivated areas (2)	2017	China, Tanzania	2,358,906 (280.59)	12,874 (3.51)	0.5%
BO1	Illumina	Plant (49)	Herbaria (49)	2018	France, Great Britain, Mauritius, Netherlands, South Africa, USA	6,621,986 (542.79)	2,689 (0.534)	0.04%
MP1	Illumina	Plant (3)	Cultivated (1), uncultivated areas (2)	2018	Burkina Faso	1,459,838 (109.67)	27,852 (4.71)	1.9%
MU-1	Illumina	Plant (3)	Cultivated areas (3)	2018	Comoro Islands, Madagascar, Mayotte Island, Reunion Island	1,945,582 (191.82)	1,005 (0.55)	0.05%
SO1	Illumina	Plant (5)	Cultivated areas (5)	2018	Burkina Faso	2,955,138 (400.19)	4,873 (1.37)	0.2%
SP	Illumina	Plant (4)	Cultivated (2), uncultivated areas (2)	2018	France, USA	2,063,144 (268.32)	108,224 (17.66)	5.2%
GMN	Illumina	Plant (9)	Cultivated areas (9)	2019	Argentina, Burkina Faso, France, Ivory Coast, USA	5,401,063 (717.84)	211,249 (31.39)	3.9%
NN2	Illumina	Plant (26)	Cultivated (4), uncultivated areas (22)	2019	France	1,667,474 (217.51)	150,318 (24.11)	9.0%
TT	Illumina	Plant (27)	Uncultivated areas (27)	2019	France	6,000,603 (818.14)	87,651 (15.19)	1.5%

According to the most up-to-date taxonomic affiliations of viruses (October 2020), 189 viral families are currently established (Walker et al. 2021). Virus sequences belonging to 85 of these 189 families were identified from 17,904 contigs (>500 nt in length) retrieved from the 24 VANA studies investigated herein (Table 3). Viruses in these families are predicted to display a broad range of genomic features, structural characteristics, and natural hosts (algae, amoeba, archaea, bacteria, fungi, invertebrates, plants, and vertebrates; Table 3).

Among the viruses of the 85 viral families represented in the outputs of the 24 VANA studies, 24 are known to contain plant-infecting viruses and the remainder either algae-, amoeba-, archaea-, bacteria-, fungi-, invertebrates-, or vertebrates-infecting viruses (Table 3). While in some families, viruses are known to infect a specific host phylum, others contain viruses found in a wide range of host phyla. For instance, whereas members of families such as *Alphasatellitidae* and *Geminiviridae* are only known to infect plants, those of the families *Reoviridae* and *Metaviridae* are known to infect fungi, invertebrates, plants, and vertebrates (Table 3).

The viruses of the 85 viral families detected in the VANA-based studies belonged to five different classes based on their type of genome: dsDNA viruses (23 families), dsRNA viruses (11 families), ssDNA viruses (11 families), positive-sense ssRNA viruses (32 families), and negative-sense ssRNA viruses (8 families). These classes cover a broad range of genome structures including circular genomes, linear genomes, and segmented genomes (Table 3).

Importantly, capsidless RNA viruses of fungi, oomycetes and plants (Dolja and Koonin 2012; Fermin 2018; Krupovic and Koonin 2017) were detected in numerous VANA-based studies, including amalgaviruses for which putative capsid proteins were reported (Krupovic et al. 2015) but all attempts to visualize virus particles have so far failed (Martin et al. 2011; Sabanadzovic et al. 2009, 2010). These capsidless RNA viruses were classified in the families *Amalgaviridae* (found in 10 of the 24 VANA studies considered herein), *Deltaflexiviridae* (1/24), *Endornaviridae* (12/24), *Hypoviridae* (2/24), *Narnaviridae* (5/24), *Mitoviridae* (3/24), and *Polymycoviridae* (4/24) (Ma et al. 2019; Maclot et al. 2020). Two nonmutually exclusive possibilities may explain the good performance of the VANA approach to detect viruses with no known particles. Genomes of capsidless viruses may be encapsidated in trans into the capsids of "normal" viruses (Das et al. 2021) and, on the other hand, capsidless viral genomes may be externalized via extracellular vesicles (EV) (Kerviel et al. 2021). The EV mechanism appears to be widespread in all three domains of life, including in the domain archaea (Gaudin et al. 2014). This demonstrates that the VANA approach is not limited to virion-producing agents and is capable of detecting many capsidless viruses. Nevertheless, it must be noted that none of the assessed VANA-based studies resulted in detection of any viroids using BLASTn searches.

Discovery and characterization of novel viruses

As of January 2022, the VANA approach has been used to characterize the viral diversity associated with plant and insect samples. Some of these viruses have been assigned to existing viral genera but others were so divergent from members of known genera that it has been necessary to either create new genera to accommodate them or to leave them unassigned for the time being.

Given that the VANA approach generally yields only partial genome sequences due to low or uneven genome coverage (in most cases <20% of the genome), accurate taxonomic assignation of reads and contigs remains challenging. This is especially true for reads <100 nts derived from divergent viral lineages as highlighted by the above described analysis. Furthermore, the identification challenge is also made harder due to the biased distribution in GenBank toward viral accessions from crop plants.

To circumvent these issues, several ad hoc OTU based classification systems have been used to operationally classify metagenomics-acquired viral sequences for downstream analyses (Bernardo et al. 2018; Guo et al. 2021; Kieft et al. 2020; Lefebvre et al. 2019; Ren et al. 2017). For example, Bernardo et al (2018) used a three-step classification system for tentatively assigning VANA OTUs to known plant virus families. The first step tentatively assigned related groups of OTUs to known plant virus families using BLASTn and BLASTx estimates of pairwise sequence similarity (using an E-value threshold of <0.001). The second step involved alignment of protein sequences coded by OTUs with homologous GenBank virus accessions, followed by the generation of maximum likelihood phylogenetic trees from these alignments. These trees were then used to determine whether the VANA-derived protein sequences nested within clades containing sequences from the candidate family or separately branched basal to these previously assigned sequences. Finally, in a third step, OTUs were classified as members of known species when they shared >75% pairwise amino acid sequence identity with previously classified sequences within recognized species. OTUs were considered belonging to potentially novel species within a family when they phylogenetically clustered within the family but shared <75% amino acid sequence identity with any other sequences previously assigned to a

TABLE 2
Impact of contig length on the performance of BLASTx

			Virus family		Virus genus		Virus species	
Contig length (bp)	Number of contigs[a]	Number of BLASTx matches[b]	Similar BLASTx matches[c]	Different BLASTx matches[d]	Similar BLASTx matches	Different BLASTx matches	Similar BLASTx matches	Different BLASTx matches
1,000	160	155 (96.9%)	154 (99.4%)	1 (0.6%)	154 (99.4%)	1 (0.6%)	101 (65.2%)	54 (34.8%)
500	333	305 (91.6%)	296 (97.0)	9 (3.0%)	284 (93.1%)	21 (6.9%)	167 (54.8%)	138 (45.2%)
200	856	672 (78.5%)	658 (97.9%)	14 (2.1%)	609 (90.6%)	63 (9.4%)	332 (49.4%)	340 (50.6%)
100	1,730	897 (51.8%)	877 (97.8%)	20 (2.2%)	779 (86.7%)	118 (13.3%)	368 (41.0%)	530 (59.0%)
50	3,481	7 (0.2%)	7 (100%)	0	7 (100%)	0	5 (71.4%)	2 (28.6%)

[a] Number of subfragments, obtained after the fragmentation of the initial long sequence, that were submitted to BLASTx.
[b] Number (and percentage) of subfragments for which BLASTx returned a match (with a e-value < 0.001).
[c] Number (and percentage) of similar BLASTx taxonomic assignment, at the family level, between the long initial sequence and its corresponding short trimmed sequences (from 50 nt to 1,000 nt).
[d] Number (and percentage) of different BLASTx taxonomic assignment, at the family level, between the long initial contig and its corresponding short trimmed sequences.

TABLE 3
List of the virus families for which contigs >500 nt were detected by the virion-associated nucleic acids (VANA)-based approach[a]

Family	Genome characteristics		Host(s)						Morphology	Virion characteristics and architectural class	Envelop	Number of contigs
	Nucleic acid	Genomic molecule(s)	P	F	B	V	I	Ar	Am	Al					
Ackermannviridae	dsDNA	1L				X					Icosahedral head with tail	HK97-like fold	N	1
Adenoviridae	dsDNA	1L					X				Icosahedral	Jellyroll, double	N	82
Alphaflexiviridae	ssRNA (+)	1L	X	X							Filamentous, flexible	Phlebo NC-like	N	1,489
Alphasatellitidae	ssDNA	1C	X								Icosahedral	Jellyroll, single	N	57
Amalgaviridae	dsRNA	1L	X	X							Capsidless	None	NA	189
Asfarviridae	dsDNA	1L					X				Icosahedral	Jellyroll, double	Y	1
Astroviridae	ssRNA (+)	1L					X				Icosahedral	Jellyroll, single	N	6
Autographiviridae	dsDNA	1L				X					Icosahedral head with tail	HK97-like fold	N	10
Bacilladnaviridae	ssDNA (dsDNA)	1C								X	Icosahedral	Jellyroll, single	N	1
Baculoviridae	dsDNA	1C					X				Helical nucleocapsid	NA	Y	40
Barnaviridae	ssRNA (+)	1L			X						Bacilliform	Jellyroll, single	N	2
Benyviridae	ssRNA (+)	4-5L - multipartite	X								Filamentous, rigid	Alpha helix-bundle, TMV-like	N	6
Betaflexiviridae	ssRNA (+)	1L	X	X							Filamentous, flexible	Phlebo NC-like	N	412
Bidnaviridae	ssDNA	2L - multipartite					X				Icosahedral	Jellyroll, single	N	9
Birnaviridae	dsRNA	2L - segmented					X	X			Icosahedral	Jellyroll, single	N	41
Bromoviridae	ssRNA (+)	3L - multipartite	X								Icosahedral/bacilliform	Jellyroll, single	N	319
Caliciviridae	ssRNA (+)	1L					X				Icosahedral	Jellyroll, single	N	1
Carmotetraviridae	ssRNA (+)	1L					X				Icosahedral	Jellyroll, single	N	5
Caulimoviridae	dsDNA	1C	X				X				Icosahedral/bacilliform	Alpha-helical (SCAN domain), retro-like	N	952
Chrysoviridae	dsRNA	4L - multipartite	X	X			X				Icosahedral	Reo-like	N	188
Circoviridae	ssDNA	1C					X				Icosahedral	Jellyroll, single	N	230
Closteroviridae	ssRNA (+)	1-3L - multipartite	X								Filamentous, flexible	Phlebo NC-like	N	350
Deltaflexiviridae	ssRNA (+)	1L	X	X	X						Probably capsidless	None	NA	12
Demerecviridae	dsDNA	1L		X							Icosahedral head with tail	HK97-like fold	N	2
Dicistroviridae	ssRNA (+)	1L					X				Icosahedral	Jellyroll, single	N	53
Drexlerviridae	dsDNA	1L			X						Icosahedral head with tail	HK97-like fold	N	3
Endornaviridae	dsRNA	1L	X	X							Capsidless	None	NA	838
Flaviviridae	ssRNA (+)	1L					X				Icosahedral	Alpha-helical, basic protein	Y	13
Geminiviridae	ssDNA	1-2C - mono/multipartite	X								Icosahedral	Jellyroll, single	N	707
Genomoviridae	ssDNA	1C		X		X					Icosahedral	NA	N	949
Hepeviridae	ssRNA (+)	1L					X				Icosahedral	Jellyroll, single	N	2
Herelleviridae	dsDNA	1L			X						Icosahedral head with tail	NA	N	3
Herpesviridae	dsDNA	1L					X				Icosahedral	HK97-like fold	Y	16
Hypoviridae	dsRNA	1L		X							Capsidless	None	NA	38
Hytrosaviridae	dsDNA	1C					X				Helical nucleocapsid	Unknown, baculo-like	Y	2
Iflaviridae	ssRNA (+)	1L					X				Icosahedral	Jellyroll, single	N	144
Inoviridae	ssDNA	1C			X						Filamentous, flexible	Alpha helix	N	6
Iridoviridae	dsDNA	1L					X	X			Icosahedral	Jellyroll, double	N	15
Kitaviridae	ssRNA (+)	2-3-4L - multipartite	X								Bacilliform	NA	N	1
Marseilleviridae	dsDNA	1C								X	Icosahedral	Jellyroll, double	N	1
Medioniviridae	ssRNA (+)	1L					X				NA	NA	Y	9
Megabirnaviridae	dsRNA	2L - multipartite	X								Icosahedral	NA	N	17
Mesoniviridae	ssRNA (+)	1L					X				Spherical	Corona-like NC	Y	2
Metaviridae	ssRNA (+)	1L	X	X		X	X				Ovoidal	Alpha-helical (SCAN domain), retro-like	Y	65

(*Continued on next page*)

[a] This table is organized in terms of virus family names ordered alphabetically. The abbreviations for the hosts are as follows: P, plant; F, fungus; B, bacteria; V, vertebrate; I, invertebrate; Ar, archaea; Am, amoeba; and Al, alga. The abbreviations for the type of genomic molecules are as follows: L, linear; and C, circular. Virion features, morphology, and architectural class were obtained from Krupovic and Koonin (2017).

TABLE 3
(*Continued from previous page*)

Family	Genome characteristics		Host(s)						Morphology	Virion characteristics and architectural class	Envelop	Number of contigs
	Nucleic acid	Genomic molecule(s)	P	F	B	V	I	Ar	Am	Al					
Microviridae	ssDNA	1C			X						Icosahedral	Jellyroll, single	N	632
Mimiviridae	dsDNA	1L					X				Icosahedral	Jellyroll, double	N	3
Mitoviridae	ssRNA (+)	1L		X							Capsidless	None	NA	7
Mymonaviridae	ssRNA (–)	1L		X							Filamentous, flexible, helical nucleocapsid	Borna-like NC	Y	21
Myoviridae	dsDNA	1L			X			X			Icosahedral head with tail	HK97-like fold	N	106
Nanoviridae	ssDNA	6 or 8C - multipartite	X								Icosahedral	NA	N	103
Narnaviridae	ssRNA (+)	1L		X							Capsidless	None	NA	5
Nodaviridae	ssRNA (+)	2L - segmented				X	X				Icosahedral	Jellyroll, single	N	28
Nudiviridae	dsDNA	1C				X	X				Helical nucleocapsid	Unknown, baculo-like	Y	1
Nyamiviridae	ssRNA (–)	1L				X	X				Spherical, helical nucleocapsid	Borna-like NC	Y	2
Papillomaviridae	dsDNA	1C				X					Icosahedral	Jellyroll, single	N	18
Paramyxoviridae	ssRNA (–)	1L				X					Spherical, helical nucleocapsid	Borna-like NC	Y	1
Partitiviridae	dsRNA	2L - multipartite	X	X							Icosahedral	Reo-like (Picobirna-like)	N	1,469
Parvoviridae	ssDNA	1L				X	X				Icosahedral	Jellyroll, single	N	555
Phenuiviridae	ssRNA (–)	3-5L - segmented/ multipartite				X	X				Spherical, icosahedral arrangement of glycoproteins (except for tenuiviruses)	Phlebo NC-like	Yes (except for tenuiviruses)	128
Phycodnaviridae	dsDNA	1L								X	Icosahedral	Jellyroll, double	Y	7
Picobirnaviridae	dsRNA	2L - segmented			X	X					Icosahedral	Reo-like (Picobirna-like)	N	6
Picornaviridae	ssRNA (+)	1L				X					Icosahedral	Jellyroll, single	N	10
Pneumoviridae	ssRNA (–)	1L				X					Spherical, helical nucleocapsid	Borna-like NC	Y	1
Podoviridae	dsDNA	1L			X						Icosahedral	HK97-like fold	N	61
Polycipiviridae	ssRNA (+)	1L					X				Icosahedral	NA	N	4
Polydnaviridae	dsDNA	1C					X				Helical nucleocapsid	Unknown,baculo-like	Y	3
Polymycoviridae	dsRNA	4, 5 or 8L - segmented		X							Capsidless	None	NA	24
Polyomaviridae	dsDNA	1C				X					Icosahedral	Jellyroll, single	N	8
Potyviridae	ssRNA (+)	1-2L - multipartite	X								Filamentous, flexible	Phlebo NC-like	N	1,776
Poxviridae	dsDNA	1L				X	X				Brick-shaped	Jellyroll, double	Y	5
Qinviridae	ssRNA (–)	2L - segmented					X				NA	NA	NA	2
Reoviridae	dsRNA	9-12L - segmented	X	X		X	X				Icosahedral, double-layered	Reo-like	N	246
Retroviridae	ssRNA (+)	1L				X					Spherical to pleomorphic ('fullerene-cone')	Alpha-helical (SCAN domain), retro-like	Y	89
Rhabdoviridae	ssRNA (–)	1-2L - mono/multipartite	X			X	X				Bullet-shaped, helical nucleocapsid	Borna-like NC	Y	30
Secoviridae	ssRNA (+)	1-2L - mono/multipartite	X								Icosahedral	Jellyroll, single	N	74
Sinhaliviridae	ssRNA (+)	1L					X				Icosahedral	Jellyroll, single	NA	64
Siphoviridae	dsDNA	1L			X			X			Icosahedral head with tail	HK97-like fold	N	578
Smacoviridae	ssDNA	1C				X	X				Probably icosahedral	NA	NA	2
Solemoviridae	ssRNA (+)	1L	X								Icosahedral	Jellyroll, single	N	691
Togaviridae	ssRNA (+)	1L				X	X				Icosahedral	Chymotrypsin-like protease (internal)	Y	4
Tombusviridae	ssRNA (+)	1-2L - mono/multipartite	X								Icosahedral	Jellyroll, single	N	81
Totiviridae	dsRNA	1L		X							Icosahedral	Reo-like	N	567
Tymoviridae	ssRNA (+)	1L	X								Icosahedral	Jellyroll, single	N	160
Virgaviridae	ssRNA (+)	1-3L - mono/multipartite	X								Filamentous, rigid	Alpha helix-bundle, TMV-like	N	163
Yueviridae	ssRNA (–)	1-2L - segmented					X				NA	NA	NA	1

recognized species. With this OTU-based classification scheme, 94 OTUs representing potentially novel species were identified from amongst 1,725 geo-referenced plant samples. These 94 OTUs were assigned to 19 plant virus families and four recognized (but unassigned to a family) virus genera. Notably, of these 94 OTUs, 45 encoded protein sequences that shared <50% identity with those of previously assigned members of known families plausibly represented novel genera within 16 different virus families.

Another ad hoc OTU-based classification system, called VirAnnot (Lefebvre et al. 2019), uses a clustering approach with an OTU-demarcation criterion based on degrees of similarity shared by sequences encoding conserved viral protein domains. A 10% cutoff value was chosen for OTUs defined on the basis of the RdRp conserved motifs to approximate the degree of domain distance that encompasses the known diversity within virus families containing RdRp conserved motifs (Lefebvre et al. 2019). The VirAnnot classification system has been applied to datasets produced by two viral sequence enrichment approaches: VANA and dsRNA from complex plant pools (Ma et al. 2019). A total of 239 RdRp OTUs from dsRNA and VANA datasets were identified, overall representing 16 RNA virus families. In this investigation, the dsRNA-based approach consistently revealed a broader and more comprehensive diversity of RNA viruses as compared to the VANA approach. However, an early iteration of the VANA approach (Candresse et al. 2014) was used in this comparative study, and it would be interesting to determine whether this pattern persists when comparing the performance of the most recent iteration of the VANA approach with the dsRNA approach (François et al. 2018a).

While hundreds of metagenome-assembled genomes (MAGs) associated with plants and insects comprising the complete coding potential of the respective virus have been discovered using these and other ad hoc OTU-based classification systems and can be now incorporated into the ICTV taxonomy (Dutilh et al. 2021; Simmonds et al. 2017), 75 of these viruses (Table 4) have subsequently been subjected to more in-depth analyses, in most cases using Sanger sequencing (Adams et al. 2013; Bagayoko et al. 2021; Bernardo et al. 2018; Boukari et al. 2017; Candresse et al. 2014; Claverie et al. 2018; Dutta et al. 2014; Fontenele et al. 2020; François et al. 2014, 2019, 2021; Gallet et al. 2018; Grisoni et al. 2017; Kraberger et al. 2015; Kwibuka et al. 2021; Ma et al. 2021; Maclot et al. 2021; Nemchinov et al. 2018; Palanga et al. 2017, 2021; Reynard et al. 2021; Richards et al. 2014; Richet et al. 2019; Roumagnac et al. 2015; Scussel et al. 2019; Susi et al. 2017). Whereas 58 of these 75 viruses representing new species were found in plant samples (77%), 17 were from insect samples (23%) (Table 4). Among the 58 plant-associated species, 23 (40%) were from cultivated plants and 35 (60%) from noncultivated plants (Table 4). Collectively, the 75 novel species have been taxonomically assigned to 22 families and 32 genera (Table 4).

The class of ssDNA viruses was predominant among these 75 species with 31 members, including 14 assigned to the family *Geminiviridae* and 17 to other divergent circular replication-associated protein (Rep)-encoding single-stranded (CRESS) DNA virus lineages (Bernardo et al. 2013; Claverie et al. 2018). The discovery using metagenomics methods of these often highly divergent species prompted reevaluation of the taxonomy of the *Geminiviridae* family. This resulted in an increase from four established genera (*Begomovirus*, *Curtovirus*, *Mastrevirus*, and *Topocuvirus*) in 2014, to 14 genera in 2022 (Roumagnac et al. 2021). Although geminiviruses are notoriously associated with serious diseases of cultivated crops, many of the viruses in the newly discovered genera, such as those in the genus *Capulavirus*, were discovered in uncultivated hosts in which they do not cause obvious infection symptoms. This stresses the need for a better understanding of both the ecology of geminiviruses in natural ecosystems and the evolutionary processes at play during the adaptation and emergence of pathogenic geminiviruses in cropping settings (Claverie et al. 2018, 2019). The characterization of capulaviruses, and subsequently that they are transmitted by aphids, has also informed our understanding of the range of insect vectors capable of transmiting geminiviruses as this is the first identified aphid transmitted geminivirus group (Roumagnac et al. 2015; Ryckebusch et al. 2020; Susi et al. 2019).

The 17 new plant-associated CRESS DNA virus species identified during VANA-based studies, collectively referred to as plant-associated genomoviruses (Table 4), have been assigned to the genera *Gemycircularvirus*, *Gemykibivirus*, *Gemykolovirus*, and *Gemykroznavirus* (Fontenele et al. 2020; Varsani and Krupovic 2021). Although these plant-associated genomoviruses were isolated from plants, it remains unclear whether any of them actually infect plants. Other genomoviruses are known to infect fungi (Varsani and Krupovic 2021) and it is therefore plausible that, rather than infecting plants, these viruses infect endophytic fungal species.

The VANA approach has also been applied successfully to the discovery of insect viruses (François et al. 2014, 2018a, b, 2019, 2021). For example, a study focusing on the diversity and abundance of viruses associated with alfalfa weevils (*Hypera postica*) revealed the presence of five novel weevil viruses (François et al. 2021). Consequently, this kind of metagenomics analysis at the ecosystem level could be a tool for identifying novel entomopathogenic viral resources with potential utility as biocontrol agents. This study also revealed that, perhaps thanks to the accumulation/concentration of viral particles within plant-feeding insects, these insects could potentially carry an even higher diversity of plant viruses than the plants on which they feed (François et al. 2021; Roberts et al. 2018). Along the same line, plant viruses have been detected in predatory insects such as dragonflies that commonly prey on plant-feeding insects. (Rosario et al. 2012, 2014). Therefore, these insect predators/foragers could potentially be used as viral sampling tools in remote and difficult to access ecosystems (Neo and Tan 2017; Rosario et al. 2012).

Use of the VANA Approach in Diagnostics and Plant Quarantine Services

Given the decreasing sequencing costs and the increasing amounts of high quality viral sequence data that can be obtained from samples using approaches such as VANA, metagenomic-based virus detection strategies may eventually be much more commonly used than all other conventional PCR-based or serological assays that are presently used in plant-viruses surveillance and diagnostics programs (de Vries et al. 2021; López-Labrador et al. 2021).

Towards this goal, the VANA approach has been tested for its ability to detect known and novel viruses within a sugarcane quarantine context. A quarantine is a setting in which living plant materials are imported from outside a country and where they must be rigorously evaluated to ensure that they do not carry pathogens (Candresse et al. 2014). In this study, a new and highly divergent mastrevirus (sugarcane white streak virus, SWSV), was identified using VANA and siRNA approaches. SWSV was found in coinfections with another mastrevirus (sugarcane streak Egypt virus, SSEV) in two plants originating from Egypt. The potential interest of the VANA approach and other metagenomics approaches for routine quarantine diagnostics is emphasized by the fact that SWSV had escaped routine quarantine detection assays (Candresse et al. 2014).

Another study focusing on the potential interest of metagenomics in sugarcane germplasm collection settings revealed the presence of another unknown sugarcane-infecting mastrevirus, called sugarcane striate virus (SStrV), in four different species of *Saccharum* (*S. officinarum*, *S. barberi*, *S. spontaneum*, and *S. sinense*) and two commercial sugarcane hybrids (Boukari et al. 2017). Interestingly, these six plants currently conserved in germplasm collections from the western hemisphere (United States and Guadeloupe) were all originally sourced from Asian countries (India, Indonesia, Iran,

TABLE 4
List of 75 new viruses, initially detected using the virion-associated nucleic acids (VANA)-based approach, that were further validated, their full genome sequence or at least their complete coding regions having been fully obtained, in the majority of cases with the addition of Sanger sequencing

Virus family and genus	Virus species	Virus name	Plant host	Accession number	Reference
Alphaflexiviridae, Allexivirus	*Vanilla latent virus*	Vanilla latent virus	*Vanilla planifolia*	MF150239	(Grisoni et al. 2017)
Alphaflexiviridae, Marafivirus	Unclassified	Ambrosia asymptomatic virus 1	*Ambrosia psilostachya*	KF421905	(Dutta et al. 2014)
Alphaflexiviridae, Potexvirus	*Vanilla virus X*	Vanilla virus X	*Vanilla planifolia*	MF150240	(Grisoni et al. 2017)
Alphaflexiviridae, unclassified	Unclassified	Hypera postica associated alphaflexivirus	*Hypera postica*	MW676130	(François et al. 2021)
Alphasatellitidae, Sophoyesatellite	*Cow vetch latent alphasatellite*	Cow vetch latent virus alphasatellite 1	*Vicia cracca*	MF535455	(Gallet et al. 2018)
Alphasatellitidae, Sophoyesatellite	*Cow vetch latent alphasatellite*	Cow vetch latent virus alphasatellite 2	*Vicia cracca*	MF535456	(Gallet et al. 2018)
Amalgaviridae, Amalgavirus	Unclassified	Medicago sativa amalgavirus 1	*Medicago sativa*	MW676142	(François et al. 2021)
Benyviridae, Benyvirus	*Rice stripe necrosis virus*	Rice stripe necrosis virus	*Oryza sativa*	MW147222 MW147223	(Bagayoko et al. 2021)
Birnaviridae, Entomobirnavirus	Unclassified	Tetranychus urticae-associated entomobirnavirus	*Tetranychus urticae*	MK533149 MK533150	(François et al. 2019)
Closteroviridae, Ampelovirus	Unclassified	Manihot esculenta associated ampelovirus 1	*Manihot esculenta*	MT773584 MT773590	(Kwibuka et al. 2021)
Closteroviridae, Ampelovirus	Unclassified	Manihot esculenta associated ampelovirus 2	*Manihot esculenta*	MT773591 MT773592 MT773594 MT773596	(Kwibuka et al. 2021)
Closteroviridae, Closterovirus	Unclassified	Blackcurrant leafroll-associated virus 1	*Ribes nigrum*	MH541840	(Koloniuk et al. 2018)
Dicistroviridae, unclassified	Unclassified	Tetranychus urticae-associated dicistrovirus 1	*Tetranychus urticae*	MK533147	(François et al. 2019)
Dicistroviridae, unclassified	Unclassified	Tetranychus urticae-associated dicistrovirus 2	*Tetranychus urticae*	MK533148	(François et al. 2019)
Geminiviridae, Becurtovirus	*Exomis microphylla associated virus*	Exomis microphylla associated virus	*Exomis microphylla*	MG001960	(Claverie et al. 2018)
Geminiviridae, Capulavirus	*Euphorbia caput-medusae Latent virus*	Euphorbia caput-medusae Latent virus	*Euphorbia caput-medusae*	HF921459 HF921477 HF921460	(Bernardo et al. 2013)
Geminiviridae, Capulavirus	*Plantago lanceolata latent virus*	Plantago lanceolata latent virus	*Plantago lanceolata*	KT214389 KT214390	(Susi et al. 2017)
Geminiviridae, Capulavirus	*Alfalfa leaf curl virus*	Alfalfa leaf curl virus	*Medicago sativa*	KP732474	(Roumagnac et al. 2015)
Geminiviridae, Capulavirus	*Trifolium virus 1*	Trifolium virus 1	*Trifolium repens*	MW698813 MW698814 MW698815 MW698816 MW698817 MW698818	(Ma et al. 2021)
Geminiviridae, Capulavirus	*Trifolium virus 1*	Trifolium virus 1	*Medicago arborea*	MW698819 MW698820 MW698821	(Ma et al. 2021)
Geminiviridae, Maldovirus	*Juncus maritimus geminivirus 1*	Juncus maritimus geminivirus 1	*Juncus maritimus*	MG001958	(Claverie et al. 2018)
Geminiviridae, Mastrevirus	*Sugarcane white streak virus*	Sugarcane white streak virus	Sugarcane	KJ210622 KJ187745 KJ187746 KJ187747 KJ187748 KJ187749	(Candresse et al. 2014)

(*Continued on next page*)

TABLE 4
(*Continued from previous page*)

Virus family and genus	Virus species	Virus name	Plant host	Accession number	Reference
Geminiviridae, *Mastrevirus*	*Sugarcane striate virus*	Sugarcane striate virus	*Saccharum spontaneum*	KX352041 KX352042 KX352044 KX352047	(Boukari et al. 2017)
Geminiviridae, *Mastrevirus*	*Sugarcane striate virus*	Sugarcane striate virus	*Saccharum barberi*	KX352040 KX352045 KX352046	(Boukari et al. 2017)
Geminiviridae, *Mastrevirus*	*Sugarcane striate virus*	Sugarcane striate virus	*Saccharum officinarum*	KX352043 KX352048 KX352049 KX352050 KX352056	(Boukari et al. 2017)
Geminiviridae, *Mastrevirus*	*Sugarcane striate virus*	Sugarcane striate virus	Sugarcane	KX352051 KX352052 KX352053 KX352054 KX352055	(Boukari et al. 2017)
Geminiviridae, unclassified	*Polygala garcinii associated virus*	Polygala garcinii associated virus	*Polygala garcinii*	MG001959	(Claverie et al. 2018)
Geminiviridae, unclassified	*Limeum africanum associated virus*	Limeum africanum associated virus	*Limeum africanum*	MG001961	(Claverie et al. 2018)
Genomoviridae, *Gemycircularvirus*	*Gemycircularvirus willde1*	Plant associated genomovirus 13	*Willdenowia* sp.	MH939427	(Fontenele et al. 2020)
Genomoviridae, *Gemycircularvirus*	*Gemycircularvirus willde1*	Plant associated genomovirus 13	*Asparagus declinatus*	MH939434	(Fontenele et al. 2020)
Genomoviridae, *Gemycircularvirus*	*Gemycircularvirus sarpe1*	Plant associated genomovirus 17	*Salicornia perennis*	MH939397	(Fontenele et al. 2020)
Genomoviridae, *Gemycircularvirus*	*Gemycircularvirus dichism1*	Plant associated genomovirus 19	*Dischisma capitatum*	MH939446	(Fontenele et al. 2020)
Genomoviridae, *Gemycircularvirus*	*Gemycircularvirus lebec1*	Plant associated genomovirus 20	*Lebeckia* sp.	MH939431	(Fontenele et al. 2020)
Genomoviridae, *Gemycircularvirus*	*Gemycircularvirus aspar1*	Plant associated genomovirus 21	*Asparagus* sp.	MH939436	(Fontenele et al. 2020)
Genomoviridae, *Gemycircularvirus*	*Gemycircularvirus trilo1*	Plant associated genomovirus 22	*Hypochaeris* sp.	MH939442	(Fontenele et al. 2020)
Genomoviridae, *Gemycircularvirus*	*Gemycircularvirus trilo1*	Plant associated genomovirus 22	*Trilobium uniolae*	MH939445	(Fontenele et al. 2020)
Genomoviridae, Gemycirculavirus	*Gemycircularvirus bromas1*	Bromus-associated circular DNA virus-3	*Bromus hordeaceus*	KM510192	(Kraberger et al. 2015)
Genomoviridae, *Gemykibivirus*	*Gemykibivirus cynas1*	Plant associated genomovirus 3	*Cynodon* sp.	MH939438	(Fontenele et al. 2020)
Genomoviridae, *Gemykibivirus*	*Gemykibivirus cynas1*	Plant associated genomovirus 3	*Hypochaeris radicata*	MH939439	(Fontenele et al. 2020)
Genomoviridae, *Gemykolovirus*	*Gemykolovirus poaspe1*	Plant associated genomovirus 9	*Avena byzantina*	MH939382 MH939394 MH939398 MH939402 MH939406 MH939408 MH939409	(Fontenele et al. 2020)
Genomoviridae, *Gemykolovirus*	*Gemykolovirus poaspe1*	Plant associated genomovirus 9	*Bromus diandrus*	MH939383 MH939401 MH939407	(Fontenele et al. 2020)
Genomoviridae, *Gemykolovirus*	*Gemykolovirus poaspe1*	Plant associated genomovirus 9	*Cyclopia genistoides*	MH939390	(Fontenele et al. 2020)
Genomoviridae, *Gemykolovirus*	*Gemykolovirus poaspe1*	Plant associated genomovirus 9	*Lolium perenne*	MH939392	(Fontenele et al. 2020)
Genomoviridae, *Gemykolovirus*	*Gemykolovirus poaspe1*	Plant associated genomovirus 9	*Poaceae* sp.	MH939426	(Fontenele et al. 2020)

(*Continued on next page*)

TABLE 4
(*Continued from previous page*)

Virus family and genus	Virus species	Virus name	Plant host	Accession number	Reference
Genomoviridae, Gemykolovirus	*Gemykolovirus poaspe1*	Plant associated genomovirus 9	*Avena sativa*	MH939428 MH939429 MH939430	(Fontenele et al. 2020)
Genomoviridae, Gemykolovirus	*Gemykolovirus poaspe1*	Plant associated genomovirus 9	*Hordeum vulgare*	MH939437	(Fontenele et al. 2020)
Genomoviridae, Gemykroznavirus	*Gemykroznavirus solas1*	Plant associated genomovirus 23	*Ehrharta longiflora*	MH939440	(Fontenele et al. 2020)
Genomoviridae, Gemykroznavirus	*Gemykroznavirus solas1*	Plant associated genomovirus 23	*Raphanus* sp.	MH939449 MH939450	(Fontenele et al. 2020)
Genomoviridae, Gemykroznavirus	*Gemykroznavirus poaspe1*	Plant associated genomovirus 24	*Poaceae* sp.	MH939435	(Fontenele et al. 2020)
Genomoviridae, Gemykroznavirus	*Gemykroznavirus poaspe1*	Plant associated genomovirus 24	*Salvia africana coerulea*	MH939444	(Fontenele et al. 2020)
Genomoviridae, unclassified	Unclassified	Plant associated genomovirus 14	*Hordeum murinum*	MH939452	(Fontenele et al. 2020)
Genomoviridae, unclassified	Unclassified	Plant associated genomovirus 16	*Phalaris minor*	MH939396	(Fontenele et al. 2020)
Genomoviridae, unclassified	Unclassified	Plant associated genomovirus 18	*Avena sativa*	MH939432 MH939447 MH939448	(Fontenele et al. 2020)
Genomoviridae, unclassified	Unclassified	Plant associated genomovirus 18	*Raphanus* sp.	MH939451	(Fontenele et al. 2020)
Genomoviridae, unclassified	Unclassified	Plant associated genomovirus 6	*Vicia faba*	MH939453	(Fontenele et al. 2020)
Genomoviridae, unclassified	Unclassified	Plant associated genomovirus 8	*Bromus diandrus*	MH939381 MH939404	(Fontenele et al. 2020)
Genomoviridae, unclassified	Unclassified	Plant associated genomovirus 8	*Avena byzantina*	MH939389 MH939393 MH939395	(Fontenele et al. 2020)
Genomoviridae, unclassified	Unclassified	Plant associated genomovirus 8	*Helichrysum revolutum*	MH939391	(Fontenele et al. 2020)
Genomoviridae, unclassified	Unclassified	Plant associated genomovirus 8	*Avena fatua*	MH939399	(Fontenele et al. 2020)
Genomoviridae, unclassified	Unclassified	Plant associated genomovirus 8	*Asparagus rubicundus*	MH939400	(Fontenele et al. 2020)
Genomoviridae, unclassified	Unclassified	Plant associated genomovirus 8	*Sarcocornia perennis*	MH939403	(Fontenele et al. 2020)
Genomoviridae, unclassified	Unclassified	Plant associated genomovirus 8	*Manulea altissima*	MH939405	(Fontenele et al. 2020)
Genomoviridae, unclassified	Unclassified	Plant associated genomovirus 8	*Asparagus declinatus*	MH939433	(Fontenele et al. 2020)
Genomoviridae, unclassified	Unclassified	Plant associated genomovirus 8	*Emex australis*	MH939441	(Fontenele et al. 2020)
Genomoviridae, unclassified	Unclassified	Plant associated genomovirus 8	*Poaceae* sp.	MH939443	(Fontenele et al. 2020)
Iflaviridae, Iflavirus	Unclassified	Hypera postica associated iflavirus 1	*Hypera postica*	MW676131	(François et al. 2021)
Iflaviridae, Iflavirus	Unclassified	Hypera postica associated iflavirus 2	*Hypera postica*	MW676132	(François et al. 2021)
Nanoviridae, Nanovirus	*Cow vetch latent virus*	Cow vetch latent virus	*Vicia cracca*	GCA_004117295	(Gallet et al. 2018)
Narnaviridae, Narnavirus	Unclassified	Tetranychus urticae-associated narnavirus	*Tetranychus urticae*	MK533151	(François et al. 2019)
Partitiviridae, Alphapartitivirus	Unclassified	Medicago sativa alphapartitivirus 1	*Hypera postica*	MW676139 MW676140	(François et al. 2021)
Parvoviridae, Ambidensovirus	Unclassified	Tetranychus urticae-associated ambidensovirus	*Tetranychus urticae*	MK543949	(François et al. 2019)
Parvoviridae, Iteradensovirus	Unclassified	Hordeum marinum Itera-like densovirus	*Hordeum marinum*	KM576800	(François et al. 2014)
Permutotetraviridae, unclassified	Unclassified	Hypera postica associated permutotetravirus	*Hypera postica*	MW676133	(François et al. 2021)

(*Continued on next page*)

TABLE 4
(*Continued from previous page*)

Virus family and genus	Virus species	Virus name	Plant host	Accession number	Reference
Picornavirales, unclassified	Unclassified	Aphis glycines virus 1	*Tetranychus urticae*	MK533146	(François et al. 2019)
Picornavirales, unclassified	Unclassified	Tetranychus urticae-associated picorna-like virus 1	*Tetranychus urticae*	MK533157	(François et al. 2019)
Picornavirales, unclassified	Unclassified	Tetranychus urticae-associated picorna-like virus 1	*Tetranychus urticae*	MK533158	(François et al. 2019)
Potyviridae, Ipomovirus	*Cassava brown streak virus*	Cassava brown streak virus	*Manihot esculenta*	MK103392 MK103393	(Scussel et al. 2019)
Potyviridae, Ipomovirus	*Ugandan cassava brown streak virus*	Ugandan cassava brown streak virus	*Manihot esculenta*	MK103391	(Scussel et al. 2019)
Secoviridae, Cheravirus.	*Arracacha virus B*	Arracacha virus B	*Oxalis tuberosa*	JQ437415 JQ581051	(Adams et al. 2013)
Secoviridae, Nepovirus	*Potato black ringspot virus*	Potato black ringspot virus	*Arracacia* spp.	KC832889 KC832891	(Richards et al. 2014)
Secoviridae, Nepovirus	*Potato black ringspot virus*	Potato black ringspot virus	*Arracacia* spp.	KC832888 KC832895	(Richards et al. 2014)
Secoviridae, Nepovirus	Unclassified	Poaceae Liege nepovirus A	*Poa trivialis*	MW289235 MW289236	(Maclot et al. 2021)
Secoviridae, unclassified	Unclassified	Poaceae Liege virus 1	*Poa trivialis*	MW289237	(Maclot et al. 2021)
Sinhaliviridae, Sinaivirus	Unclassified	Hypera postica associated sinaivirus	*Hypera postica*	MW676134	(François et al. 2021)
Solemoviridae, Polerovirus	Unclassified	Cowpea polerovirus 1	*Vigna unguiculata*	KY364846	(Palanga et al. 2017)
Solemoviridae, Polerovirus	Unclassified	Cowpea polerovirus 2	*Vigna unguiculata*	KY364847	(Palanga et al. 2017)
Solemoviridae, Sobemovirus	Unclassified	Hypera postica associated sobemovirus 1	*Hypera postica*	MW676135	(François et al. 2021)
Solemoviridae, Sobemovirus	Unclassified	Hypera postica associated sobemovirus 2	*Hypera postica*	MW676136	(François et al. 2021)
Solemoviridae, Sobemovirus	Unclassified	Hypera postica associated sobemovirus 3	*Hypera postica*	MW676137	(François et al. 2021)
Tombusviridae, Luteovirus	*Bean leafroll virus*	Bean leafroll virus	*Medicago sativa*	MW676129	(François et al. 2021)
Tymoviridae, Marafivirus	*Alfalfa virus F*	Alfalfa virus F	*Medicago sativa*	MG676465	(Nemchinov et al. 2018)
Tymoviridae, Marafivirus	Unclassified	Pennisetum glaucum marafivirus	*Pennisetum glaucum*	MZ305310	(Palanga et al. 2021)
Unclassified ssDNA viruses	Unclassified	Bromus-associated circular DNA virus-1	*Bromus hordeaceus*	KM510189 KM510190	(Kraberger et al. 2015)
Unclassified ssDNA viruses	Unclassified	Bromus-associated circular DNA virus-2	*Bromus hordeaceus*	KM510191	(Kraberger et al. 2015)
Unclassified ssDNA viruses	Unclassified	Trifolium-associated circular DNA virus 1	*Trifolium resupinatum*	KP005453	(Kraberger et al. 2015)
Unclassified ssDNA viruses	Unclassified	Bromus-associated circular DNA virus 4	*Bromus hordeaceus*	KP005454	(Kraberger et al. 2015)
Unclassified ssDNA viruses	Unclassified	Stipagrostis associated virus	*Stipagrostis* sp.	MH425570	(Richet et al. 2019)
Unclassified ssDNA viruses	Unclassified	Panicum ecklonii associated virus	*Panicum ecklonii*	MH425571	(Richet et al. 2019)
Unclassified ssDNA viruses	Unclassified	Lolium perenne associated virus	*Lolium perenne*	MH425572	(Richet et al. 2019)
Unclassified ssDNA viruses	Unclassified	Holcus lanatus associated virus	*Holcus lanatus*	MH425573	(Richet et al. 2019)
Unclassified ssDNA viruses	Unclassified	Arctopus echinatus associated virus	*Arctopus echinatus*	MH425569	(Richet et al. 2019)
Virgaviridae, Hordeivirus	Unclassified	Ligustrum mosaic virus	*Ligustrum vulgare*	MW752157 MW752158 MW752159	(Reynard et al. 2021)

Malaysia, and New Guinea). This suggested that SStrV has an Asian origin. In the United States, the SStrV isolates were identified in plants from three varieties that were introduced more than six decades ago into the Miami (Florida, United States) world germplasm collection. Nevertheless, the novel mastrevirus was not detected up to now in commercially grown sugarcane in Florida (Boukari et al. 2017), which could be explained by the absence of the insect vector(s) of SStrV in this geographical location. VANA-based health monitoring systems of plants maintained in germplasm collections could be very useful for detecting latent or asymptomatic infections by unknown viruses. Consequently, this method could also help to prevent the inadvertent dissemination of these viruses in symptom-free plant material throughout the world.

However, despite its promise, the VANA approach presents distinct shortcomings to plant virome studies. While the VANA purification protocol appears to be relatively robust, this approach may not be effective to recover viruses from plants having high levels of either phenolic compounds or highly viscous polysaccharides. In addition, the VANA approach cannot be applied to diagnose all viruses and viroids with equal efficiency (Ma et al. 2019). Collectively, these limitations suggest that the VANA approach is presently not ideally suited for routine plant quarantine diagnostics and certification workflows (Kutnjak et al. 2021). The main reason for this is that the virus particle semipurification steps remain too cumbersome for rapid routine diagnostic workflows. In addition, the VANA approach that requires numerous experimental steps and is further based on PCR assays is likely to be susceptible to cross-contamination between samples. Moreover, the VANA approach generally involves sample multiplexing, which lowers the per-sample sequencing cost but reduces the per-sample sequencing read depth. This trade-off might be partly resolved by better enrichment for viral genetic material prior to sequencing.

While the sequencing depths achieved in VANA-based metagenomic studies have been sufficient for identifying novel and known viruses in large numbers of plant samples, it is very unlikely that they succeeded in revealing the full inventory of viruses that were present within these samples. Occurrence of a certain percentage of false negatives—potentially due to a lack of sequencing depth or missing reference virus genomes in the NCBI GenBank database (dark matter) (Kieft and Anantharaman 2022)—may not be the most important issue with respect to efforts aiming at a broad description of the virome and of its properties. However, it is a serious problem in a diagnostic setting where the focus is always the maximization of true positives and true negatives balanced against the minimization of false positives and false negatives.

The application of the VANA approach in routine diagnostics would require strict validation of its performance criteria: analytical sensitivity, analytical specificity (including inclusivity and exclusivity), repeatability, and reproducibility. These performance criteria must be acceptable for the intended use of any diagnostic test and must be rigorously compared with existing alternative tests. As for any other HTS-based assay, the use of the VANA approach in a diagnostic setting would also need to accommodate the rapidly changing technologies upon which diagnostic tests are based, a factor adding an additional layer of complexity to the adoption of the VANA approach in routine diagnostic screening contexts.

To circumvent these issues, recent studies have reported the successful use of HTS from total RNA (or ribodepleted total RNA) for routine diagnostics of plant viruses (Bester et al. 2021; Gaafar et al. 2021; Gauthier et al. 2022; Malapi-Wight et al. 2021; Soltani et al. 2021). However, the authors of these studies have also highlighted the need to improve standardization of these analyses before they can be routinely used for diagnostics. Overall, several review articles have analyzed in recent years the challenges and opportunities of virus metagenomics approaches for plant pest diagnosis (Massart et al. 2019; Olmos et al. 2018; Whattam et al. 2021).

Epidemiological Surveillance and Environmental Monitoring

Along with other metagenomic approaches, the VANA approach has the potential to reveal the role of plant viruses to agroecosystems at national and regional scales (Maclot et al. 2020). Four such inventories have been compiled over the last decade: in Burkina Faso, the United States, France, and Belgium (Daugrois et al. 2021; Filloux et al. 2018; Palanga et al. 2016). The overall goal of these surveys was to obtain snapshots of viral diversity associated with plant species (both cultivated and uncultivated) growing in specific territories (region, country) and/or at specified time periods. Besides improving the surveillance and monitoring of crop health, such studies can result in the identification of viruses that potentially pose an emergence risk of harmful crop pathogens.

Snapshots of virus distributions associated with crop plants were conducted in VANA-based studies of sugarcane (in Florida, U.S.A.) and cowpea (in Burkina Faso) (Filloux et al. 2018; Palanga et al. 2016). In Florida, samples from 214 sugarcane leaf samples were collected from different commercial sugarcane (*Saccharum* interspecific hybrids) fields and from other *Saccharum* and related species in two germplasm collections (Filloux et al. 2018). A novel umbravirus (*Tombusviridae*) and six known sugarcane viruses were detected, including sugarcane yellow leaf virus (SCYLV) with high prevalence (74%) in Florida in both commercial fields and germplasm collections.

In Burkina Faso, 312 leaf samples were obtained from cowpea plants in the Sudan (humid), Sudan-Sahel (sub-humid), and Sahel (dry) agro-climatic zones of the country. Nine viruses were identified in these samples, and six of these viruses were not previously reported in Burkina Faso (Palanga et al. 2016). As in the Florida study, one of the viruses (cowpea aphid-borne mosaic virus) was shown to be highly prevalent (64%) throughout the country. Furthermore, three of the detected viruses were novel, including two poleroviruses (*Solemoviridae*) for which full genome sequences were subsequently determined (Palanga et al. 2017).

The Florida and Burkina Faso VANA-based studies demonstrated the viability of systematic periodic surveillance of crop viromes across ~50,000 km^2 regions using the VANA approach. This is a first step towards regular routine monitoring of endemic and emergent crop virus distributions and diversity at regional, national, or continental scales. Nevertheless, future VANA-based epidemiosurveillance studies will likely need to be based on a larger number of samples per unit of surface area in order to improve the overall reliability of routine monitoring of plant virus diseases.

The real power of the VANA approach is that it does not need to focus exclusively on individual crop species. This holistic approach can be illustrated with a study conducted in Belgium between 2018 and 2020 on viruses infecting 17,600 plants from 24 different genera of the family *Solanaceae* (including uncultivated and cultivated species). In this study, 40 plant viruses were identified (excluding persistent mycoviruses, invertebrate viruses, environmental viruses, and unclassified viruses). Three of these 40 viruses potentially belonged to new species and 23 had never been reported before in Belgium. Out of the 70 virus−host associations observed, 30 were previously unreported in the literature (A. G. Blouin, *personal communication*). At least one virus represented a potential threat for Belgium and was further characterized (C. Temple, *personal communication*).

Smaller scale VANA-based studies can also be illuminating. For example, comparison between the plant virome within a plant quarantine station and the virome in the surrounding flora can be performed to identify potential confinement failures and risks of cross infections (Daugrois et al. 2021). Recently, an analysis of samples from sugarcane quarantined plants in France and from uncultivated *Poaceae* plants growing in areas bordering the quarantine glasshouses was undertaken (Daugrois et al. 2021). This study

revealed that, whereas viruses belonging to the same genera and families were found in and around the quarantine facility, no virus species was detected in both environments. Such routine surveillance in targeted environments that are at risk of being an entry point of foreign viruses (e.g., near quarantine stations, seaports, airports or in areas around laboratories carrying out live-virus experiments), usually termed as biosafety in containment (Beeckman and Rüdelsheim 2020), could be used to ensure the proper implementation and efficacy of virus confinement procedures in such facilities (Daugrois et al. 2021).

Plant Virus Ecology

To understand viral emergence, viral metagenomic studies (including VANA-based studies) have been conducted at different spatial scales with plants in both managed and unmanaged environments. Agroecological interfaces and other environments that have been disturbed by humans have also been targeted by these studies (Alexander et al. 2014). At interfaces between natural and human-disturbed biomes, changes in environmental factors such as the abundances of virus hosts and vector species are expected to be more abrupt than in natural environments (Alexander et al. 2014). The main results obtained in these studies were recently summarized and reported in several reviews (Claverie et al. 2018; Hasiów-Jaroszewska et al. 2021; Lefeuvre et al. 2019; Maclot et al. 2020; McLeish et al. 2021; Sommers et al. 2021).

Briefly, the VANA approach was used to address a number of virus ecology questions in several landscape-scale situations. The first two studies were conducted in France and South Africa and used the VANA approach that accounted for the spatial arrangements of plant samples and the precise environmental contexts of individual sampling sites (Bernardo et al. 2018). Specifically, these studies aimed at assessing whether plant-associated virus communities were more prevalent but less diverse in cultivated areas and whether viruses from particular known families were significantly associated with cultivated or uncultivated areas. These studies revealed the impact of agriculture on the distribution and prevalence of plant viruses at the ecosystem scale, demonstrating that virus prevalence was greater in cultivated areas in France and South Africa and that some virus families were shown to have strong associations with agriculture. On the other hand, 80 of the 94 (84%) tentative new viruses discovered, were from uncultivated plants, thus highlighting the bias of our current knowledge towards crop-infecting viruses. Furthermore, the prevalence of plant viruses was significantly higher in exotic plants than in indigenous plants in the endangered South African fynbos biome (Bernardo et al. 2018). Finally, dsRNA viruses displayed the greatest diversity across both biomes. This result was further confirmed in a dsRNA-based metagenomics study led in the Bordeaux region (France) (Ma et al. 2019). However, a lower species richness was observed overall for dsRNA viruses from cultivated sites. Since many dsRNA viruses have persistent lifestyles in unmanaged environments, and since at least some of them are associated with fungal hosts, it has been proposed that this trend might reflect the indirect impact of fungicide treatments applied to crops (Ma et al. 2019).

The third landscape-scale VANA-based study was conducted in Belgium and focused on three different *Poaceae* communities displaying different degrees of both biodiversity (in terms of grass species richness) and anthropogenic management (i.e., cereal crops, grazed pastures and mowed grasslands). A diverse virome was identified in cultivated and noncultivated *Poaceae* comprising at least 50 species from 21 genera in 16 families (F. Maclot, V. Debue, C. M. Malmstrom, D. Filloux, P. Roumagnac, M. Eck, L. Tamisier, T. Candresse, and S. Massart, *unpublished data*). As in the French and South African contexts, dsRNA viruses with persistent lifestyles belonging to the *Alphachrysovirus*, *Partitivirus*, and *Totivirus* genera represented a large fraction of this virome, i.e., 60% of the detected viruses. The virome was compared among *Poaceae* communities and species. Among the obtained results, differences in virome composition were observed between dominant and low occurrence grasses in grasslands (i.e., representing 5 to 20% of the sampled grasses). Besides the ubiquitous persistent viruses found in all wild grasses, a number of viral taxa were only detected in the dominant grasses (e.g., *Panicovirus, Polerovirus*, and *Umbravirus*) or in the low occurrence ones (e.g., *Alphaendornavirus, Potyvirus*, and *Sobemovirus*).

Numerous methodological innovations were leveraged to reveal the ecology of plant viruses in these and other VANA-based studies, and in those that used other approaches (Hasiów-Jaroszewska et al. 2021; Maclot et al. 2020). These innovations are still under active development. They should ultimately yield experimental and analytical frameworks that achieve minimally biased sampling, sequencing and taxonomic assignation of viral sequences within natural and human-impacted environments. Efforts are ongoing to comprehensively quantify and track changes in plant virus diversity and distributions in relation to plant species richness across agricultural and ecological boundaries. This should enable the parameterization and mathematical modeling of the complex interplay between the genetic richness of plant virus populations at varying taxonomic scales (from families to individual genetic sublineages within species), physical environmental factors (such as precipitation, soil chemistry, or fire cycles), and biotic environmental factors (such as host and vector species genetic richness, demographics, and distributions). These models should provide a better understanding of plant virus diversity, distributions, host ranges, and transmission modes. In the coming years, tremendous advances are expected in our ability to predictively model these complex interactions and avoid or modify the conditions under which plant virus emergence events are most likely to occur.

Acknowledgments

We thank ViralZone for granting us the use of the wonderful virion pictures.

Literature Cited

Adams, I. P., Glover, R., Souza-Richards, R., Bennett, S., Hany, U., and Boonham, N. 2013. Complete genome sequence of arracacha virus B: A novel cheravirus. Arch. Virol. 158:909-913.

Alexander, H. M., Mauck, K. E., Whitfield, A. E., Garrett, K. A., and Malmstrom, C. M. 2014. Plant-virus interactions and the agro-ecological interface. Eur. J. Plant Pathol. 138:529-547.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403-410.

Bagayoko, I., Celli, M. G., Romay, G., Poulicard, N., Pinel-Galzi, A., Julian, C., Filloux, D., Roumagnac, P., Sérémé, D., Bragard, C., and Hébrard, E. 2021. Genetic diversity of rice stripe necrosis virus and new insights into evolution of the genus *Benyvirus*. Viruses 13:737.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., and Pevzner, P. A. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19:455-477.

Beeckman, D. S. A., and Rüdelsheim, P. 2020. Biosafety and biosecurity in containment: A regulatory overview. Front. Bioeng. Biotechnol. 8:650.

Bernardo, P., Charles-Dominique, T., Barakat, M., Ortet, P., Fernandez, E., Filloux, D., Hartnady, P., Rebelo, T. A., Cousins, S. R., Mesleard, F., Cohez, D., Yavercovski, N., Varsani, A., Harkins, G. W., Peterschmitt, M., Malmstrom, C. M., Martin, D. P., and 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.

Bernardo, P., Golden, M., Akram, M., Naimuddin, Nadarajan, N., Fernandez, E., Granier, M., Rebelo, A. G., Peterschmitt, M., Martin, D. P., and Roumagnac, P. 2013. Identification and characterisation of a highly divergent geminivirus: Evolutionary and taxonomic implications. Virus Res. 177:35-45.

Bester, R., Cook, G., Breytenbach, J. H. J., Steyn, C., De Bruyn, R., and Maree, H. J. 2021. Towards the validation of high-throughput sequencing (HTS)

for routine plant virus diagnostics: Measurement of variation linked to HTS detection of citrus viruses and viroids. Virol. J. 18:61.
Boukari, W., Alcala-Briseno, R. I., Kraberger, S., Fernandez, E., Filloux, D., Daugrois, J. H., Comstock, J. C., Lett, J. M., Martin, D. P., Varsani, A., Roumagnac, P., Polston, J. E., and Rott, P. C. 2017. Occurrence of a novel mastrevirus in sugarcane germplasm collections in Florida, Guadeloupe and Reunion. Virol. J. 14:146.
Breitbart, M., Hewson, I., Felts, B., Mahaffy, J. M., Nulton, J., Salamon, P., and Rohwer, F. 2003. Metagenomic analyses of an uncultured viral community from human feces. J. Bacteriol. 185:6220-6223.
Buchfink, B., Reuter, K., and Drost, H.-G. 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat. Methods 18:366-368.
Call, L., Nayfach, S., and Kyrpides, N. C. 2021. Illuminating the virosphere through global metagenomics. Annu. Rev. Biomed. Data Sci. 4: 369-391.
Camarillo-Guerrero, L. F., Almeida, A., Rangel-Pineros, G., Finn, R. D., and Lawley, T. D. 2021. Massive expansion of human gut bacteriophage diversity. Cell 184:1098-1109.
Candresse, T., Filloux, D., Muhire, B., Julian, C., Galzi, S., Fort, G., Bernardo, P., Daugrois, J. H., Fernandez, E., Martin, D. P., Varsani, A., and Roumagnac, P. 2014. Appearances can be deceptive: Revealing a hidden viral infection with deep sequencing in a plant quarantine context. PLoS One 9:e102945.
Claverie, S., Bernardo, P., Kraberger, S., Hartnady, P., Lefeuvre, P., Lett, J. M., Galzi, S., Filloux, D., Harkins, G. W., Varsani, A., Martin, D. P., and Roumagnac, P. 2018. From spatial metagenomics to molecular characterization of plant viruses: A geminivirus case study. Adv. Virus Res. 101: 55-83.
Claverie, S., Ouattara, A., Hoareau, M., Filloux, D., Varsani, A., Roumagnac, P., Martin, D. P., Lett, J.-M., and Lefeuvre, P. 2019. Exploring the diversity of Poaceae-infecting mastreviruses on Reunion Island using a viral metagenomics-based approach. Sci. Rep. 9:12716.
Das, S., Alam, M. M., Zhang, R., Hisano, S., and Suzuki, N. 2021. Proof of concept of the Yadokari nature: A capsidless replicase-encoding but replication-dependent positive-sense single-stranded RNA virus hosted by an unrelated double-stranded RNA virus. J. Virol. 95:e0046721.
Daugrois, J. H., Filloux, D., Julian, C., Claude, L., Ferdinand, R., Fernandez, E., Fontes, H., Rott, P. C., and Roumagnac, P. 2021. Comparison of the virome of quarantined sugarcane varieties and the virome of grasses growing near the quarantine station. Viruses 13:922.
Dávila-Ramos, S., Castelán-Sánchez, H. G., Martínez-Ávila, L., Sánchez-Carbente, M. d. R., Peralta, R., Hernández-Mendoza, A., Dobson, A. D. W., Gonzalez, R. A., Pastor, N., and Batista-García, R. A. 2019. A review on viral metagenomics in extreme environments. Front. Microbiol. 10:2403.
de Vries, J. J. C., Brown, J. R., Couto, N., Beer, M., Le Mercier, P., Sidorov, I., Papa, A., Fischer, N., Oude Munnink, B. B., Rodriquez, C., Zaheri, M., Sayiner, A., Hönemann, M., Pérez-Cataluña, A., Carbo, E. C., Bachofen, C., Kubacki, J., Schmitz, D., Tsioka, K., Matamoros, S., Höper, D., Hernandez, M., Puchhammer-Stöckl, E., Lebrand, A., Huber, M., Simmonds, P., Claas, E. C. J., and López-Labrador, F. X. 2021. Recommendations for the introduction of metagenomic next-generation sequencing in clinical virology, part II: Bioinformatic analysis and reporting. J. Clin. Virol. 138:104812.
Dodds, J. A., Morris, T. J., and Jordan, R. L. 1984. Plant viral double-stranded RNA. Annu. Rev. Phytopathol. 22:151-168.
Dolja, V. V., and Koonin, E. V. 2012. Capsid-less RNA viruses. eLS.
Dutilh, B. E., Varsani, A., Tong, Y., Simmonds, P., Sabanadzovic, S., Rubino, L., Roux, S., Muñoz, A. R., Lood, C., Lefkowitz, E. J., Kuhn, J. H., Krupovic, M., Edwards, R. A., Brister, J. R., Adriaenssens, E. M., and Sullivan, M. B. 2021. Perspective on taxonomic classification of uncultivated viruses. Curr. Opin. Virol. 51:207-215.
Dutta, M., Sokhandan Bashir, N., Palmer, M. W., and Melcher, U. 2014. Genomic characterization of Ambrosia asymptomatic virus 1 and evidence of other *Tymovirales* members in the Oklahoma tallgrass prairie revealed by sequence analysis. Arch. Virol. 159:1755-1764.
Edgar, R. C., Taylor, J., Lin, V., Altman, T., Barbera, P., Meleshko, D., Lohr, D., Novakovsky, G., Buchfink, B., Al-Shayeb, B., Banfield, J. F., de la Peña, M., Korobeynikov, A., Chikhi, R., and Babaian, A. 2022. Petabase-scale sequence alignment catalyses viral discovery. Nature 602:142-147.
Fermin, G. 2018. Virion structure, genome organization, and taxonomy of viruses. Viruses 2018:17-54.
Filloux, D., Fernandez, E., Comstock, J. C., Mollov, D., Roumagnac, P., and Rott, P. 2018. Viral metagenomic-based screening of sugarcane from Florida reveals accurrence of six sugarcane-infecting viruses and high prevalence of Sugarcane yellow leaf virus. Plant Dis. 102:2317-2323.
Fontenele, R. S., Roumagnac, P., Richet, C., Kraberger, S., Stainton, D., Aleamotu'a, M., Filloux, D., Bernardo, P., Harkins, G. W., McCarthy, J., Charles, L. S., Lamas, N. S., Abreu, E. F. M., Abreu, R. A., Batista, G. B., Lacerda, A. L. M., Salywon, A., Wojciechowski, M. F., Majure, L. C., Martin, D. P., Ribeiro, S. G., Lefeuvre, P., and Varsani, A. 2020. Diverse genomoviruses representing twenty-nine species identified associated with plants. Arch. Virol. 165:2891-2901.
François, S., Antoine-Lorquin, A., Kulikowski, M., Frayssinet, M., Filloux, D., Fernandez, E., Roumagnac, P., Froissart, R., and Ogliastro, M. 2021. Characterisation of the viral community associated with the alfalfa weevil (*Hypera postica*) and its host plant, alfalfa (*Medicago sativa*). Viruses 13:791.
François, S., Bernardo, P., Filloux, D., Roumagnac, P., Yaverkovski, N., Froissart, R., and Ogliastro, M. 2014. A novel itera-like densovirus isolated by viral metagenomics from the sea barley *Hordeum marinum*. Genome Announc. 2:e01196-01114.
François, S., Filloux, D., Fernandez, E., Ogliastro, M., and Roumagnac, P. 2018a. Viral metagenomics approaches for high-resolution screening of multiplexed arthropod and plant viral communities. Methods Mol. Biol. 1746:77-95.
François, S., Filloux, D., Frayssinet, M., Roumagnac, P., Martin, D. P., Ogliastro, M., and Froissart, R. 2018b. Increase in taxonomic assignment efficiency of viral reads in metagenomic studies. Virus Res. 244:230-234.
François, S., Mutuel, D., Duncan, A. B., Rodrigues, L. R., Danzelle, C., Lefevre, S., Santos, I., Frayssinet, M., Fernandez, E., Filloux, D., Roumagnac, P., Froissart, R., and Ogliastro, M. 2019. A new prevalent densovirus discovered in Acari. Insight from metagenomics in viral communities associated with two-spotted mite (*Tetranychus urticae*) populations. Viruses 11:233.
Gaafar, Y. Z. A., Westenberg, M., Botermans, M., László, K., De Jonghe, K., Foucart, Y., Ferretti, L., Kutnjak, D., Pecman, A., Mehle, N., Kreuze, J., Muller, G., Vakirlis, N., Beris, D., Varveri, C., and Ziebell, H. 2021. Interlaboratory comparison study on ribodepleted total RNA high-throughput sequencing for plant virus diagnostics and bioinformatic competence. Pathogens 10:1174.
Gallet, R., Kraberger, S., Filloux, D., Galzi, S., Fontes, H., Martin, D. P., Varsani, A., and Roumagnac, P. 2018. Nanovirus-alphasatellite complex identified in *Vicia cracca* in the Rhône delta region of France. Arch. Virol. 163:695-700.
Gaudin, M., Krupovic, M., Marguet, E., Gauliard, E., Cvirkaite-Krupovic, V., Le Cam, E., Oberto, J., and Forterre, P. 2014. Extracellular membrane vesicles harbouring viral genomes. Environ. Microbiol. 16:1167-1175.
Gauthier, M.-E. A., Lelwala, R. V., Elliott, C. E., Windell, C., Fiorito, S., Dinsdale, A., Whattam, M., Pattemore, J., and Barrero, R. A. 2022. Side-by-side comparison of post-entry quarantine and high throughput sequencing methods for virus and viroid diagnosis. Biology 11:263.
Greninger, A. L. 2018. A decade of RNA virus metagenomics is (not) enough. Virus Res. 244:218-229.
Grisoni, M., Marais, A., Filloux, D., Saison, A., Faure, C., Julian, C., Theil, S., Contreras, S., Teycheney, P. Y., Roumagnac, P., and Candresse, T. 2017. Two novel *Alphaflexiviridae* members revealed by deep sequencing of the *Vanilla* (*Orchidaceae*) virome. Arch. Virol. 162:3855-3861.
Guo, J., Bolduc, B., Zayed, A. A., Varsani, A., Dominguez-Huerta, G., Delmont, T. O., Pratama, A. A., Gazitúa, M. C., Vik, D., Sullivan, M. B., and Roux, S. 2021. VirSorter2: A multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. Microbiome 9:37.
Hasiów-Jaroszewska, B., Boezen, D., and Zwart, M. P. 2021. Metagenomic studies of viruses in weeds and wild plants: A powerful approach to characterise variable virus communities. Viruses 13:1939.
Kerviel, A., Zhang, M., and Altan-Bonnet, N. 2021. A new infectious unit: Extracellular vesicles carrying virus populations. Annu. Rev. Cell Dev. Biol. 37:171-197.
Kieft, K., and Anantharaman, K. 2022. Virus genomics: What is being overlooked? Curr. Opin. Virol. 53:101200.
Kieft, K., Zhou, Z., and Anantharaman, K. 2020. VIBRANT: Automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. Microbiome 8:90.
Koloniuk, I., Thekke-Veetil, T., Reynard, J.-S., Mavrič Pleško, I., Přibylová, J., Brodard, J., Kellenberger, I., Sarkisova, T., Špak, J., Lamovšek, J., Massart, S., Ho, T., Postman, J. D., and Tzanetakis, I. E. 2018. Molecular characterization of divergent closterovirus isolates infecting ribes species. Viruses 10:369.
Kraberger, S., Farkas, K., Bernardo, P., Booker, C., Arguello-Astorga, G. R., Mesleard, F., Martin, D. P., Roumagnac, P., and Varsani, A. 2015. Identification of novel *Bromus*- and *Trifolium*-associated circular DNA viruses. Arch. Virol. 160:1303-1311.
Krishnamurthy, S. R., and Wang, D. 2017. Origins and challenges of viral dark matter. Virus Res. 239:136-142.
Krupovic, M., Dolja, V. V., and Koonin, E. V. 2015. Plant viruses of the *Amalgaviridae* family evolved via recombination between viruses with double-stranded and negative-strand RNA genomes. Biol. Direct 10:12.
Krupovic, M., and Koonin, E. V. 2017. Multiple origins of viral capsid proteins from cellular ancestors. Proc. Natl. Acad. Sci. 114:E2401.
Kutnjak, D., Tamisier, L., Adams, I., Boonham, N., Candresse, T., Chiumenti, M., De Jonghe, K., Kreuze, J. F., Lefebvre, M., Silva, G., Malapi-Wight, M., Margaria, P., Mavrič Pleško, I., McGreig, S., Miozzi, L., Remenant, B., Reynard, J. S., Rollin, J., Rott, M., Schumpp, O., Massart, S., and Haegeman,

A. 2021. A primer on the analysis of high-throughput sequencing data for detection of plant viruses. Microorganisms 9:841.
Kwibuka, Y., Bisimwa, E., Blouin, A. G., Bragard, C., Candresse, T., Faure, C., Filloux, D., Lett, J.-M., Maclot, F., Marais, A., Ravelomanantsoa, S., Shakir, S., Vanderschuren, H., and Massart, S. 2021. Novel ampeloviruses infecting cassava in Central Africa and the South-West Indian Ocean Islands. Viruses 13:8411030.
Langmead, B. 2010. Aligning short sequencing reads with bowtie. Curr. Protoc. Bioinform. 32:11.17.1-11.17.14.
Lefebvre, M., Theil, S., Ma, Y., and 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.
Lefeuvre, P., Martin, D. P., Elena, S. F., Shepherd, D. N., Roumagnac, P., and Varsani, A. 2019. Evolution and ecology of plant viruses. Nat. Rev. Microbiol. 17:632-644.
López-Labrador, F. X., Brown, J. R., Fischer, N., Harvala, H., Van Boheemen, S., Cinek, O., Sayiner, A., Madsen, T. V., Auvinen, E., Kufner, V., Huber, M., Rodriguez, C., Jonges, M., Hönemann, M., Susi, P., Sousa, H., Klapper, P. E., Pérez-Cataluña, A., Hernandez, M., Molenkamp, R., der Hoek, L. V., Schuurman, R., Couto, N., Leuzinger, K., Simmonds, P., Beer, M., Höper, D., Kamminga, S., Feltkamp, M. C. W., Rodríguez-Díaz, J., Keyaerts, E., Nielsen, X. C., Puchhammer-Stöckl, E., Kroes, A. C. M., Buesa, J., Breuer, J., Claas, E. C. J., and de Vries, J. J. C. 2021. Recommendations for the introduction of metagenomic high-throughput sequencing in clinical virology, part I: Wet lab procedure. J. Clin. Virol. 134:104691.
Ma, Y., Marais, A., Lefebvre, M., Theil, S., Svanella-Dumas, L., Faure, C., and 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.
Ma, Y., Svanella-Dumas, L., Julian, C., Galzi, S., Fernandez, E., Yvon, M., Pirolles, E., Lefebvre, M., Filloux, D., Roumagnac, P., and Candresse, T. 2021. Genome characterization and diversity of trifolium virus 1: Identification of a novel legume-infecting capulavirus. Arch. Virol. 166:2573-2578.
Maclot, F., Candresse, T., Filloux, D., Malmstrom, C. M., Roumagnac, P., van der Vlugt, R., and Massart, S. 2020. Illuminating an ecological blackbox: Using high throughput sequencing to characterize the plant virome across scales. Front. Microbiol. 11:578064.
Maclot, F. J., Debue, V., Blouin, A. G., Fontdevila Pareta, N., Tamisier, L., Filloux, D., and Massart, S. 2021. Identification, molecular and biological characterization of two novel secovirids in wild grass species in Belgium. Virus Res. 298:198397.
Malapi-Wight, M., Adhikari, B., Zhou, J., Hendrickson, L., Maroon-Lango, C. J., McFarland, C., Foster, J. A., and Hurtado-Gonzales, O. P. 2021. HTS-based diagnostics of sugarcane viruses: Seasonal variation and its implications for accurate detection. Viruses 13:1627.
Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet 17:10.
Martin, R. R., Zhou, J., and Tzanetakis, I. E. 2011. Blueberry latent virus: An amalgam of the *Partitiviridae* and *Totiviridae*. Virus Res. 155:175-180.
Massart, S., Chiumenti, M., De Jonghe, K., Glover, R., Haegeman, A., Koloniuk, I., Komínek, P., Kreuze, J., Kutnjak, D., Lotos, L., Maclot, F., Maliogka, V., Maree, H. J., Olivier, T., Olmos, A., Pooggin, M. M., Reynard, J.-S., Ruiz-García, A. B., Safarova, D., Schneeberger, P. H. H., Sela, N., Turco, S., Vainio, E. J., Varallyay, E., Verdin, E., Westenberg, M., Brostaux, Y., and Candresse, T. 2019. Virus detection by high-throughput sequencing of small RNAs: Large-scale performance testing of sequence analysis strategies. Phytopathology 109:488-497.
McLeish, M. J., Fraile, A., and García-Arenal, F. 2021. Population genomics of plant viruses: The ecology and evolution of virus emergence. Phytopathology 111:32-39.
Melcher, U., Muthukumar, V., Wiley, G. B., Min, B. E., Palmer, M. W., Verchot-Lubicz, J., Ali, A., Nelson, R. S., Roe, B. A., Thapa, V., and Pierce, M. L. 2008. Evidence for novel viruses by analysis of nucleic acids in virus-like particle fractions from *Ambrosia psilostachya*. J. Virol. Methods 152:49-55.
Muthukumar, V., Melcher, U., Pierce, M., Wiley, G. B., Roe, B. A., Palmer, M. W., Thapa, V., Ali, A., and 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.
Nemchinov, L. G., François, S., Roumagnac, P., Ogliastro, M., Hammond, R. W., Mollov, D. S., and Filloux D. 2018. Characterization of alfalfa virus F, a new member of the genus Marafivirus. PLoS One 13:e0203477.
Neo, J. P. S., and Tan, B. H. 2017. The use of animals as a surveillance tool for monitoring environmental health hazards, human health hazards and bioterrorism. Vet. Microbiol. 203:40-48.
Neri, U., Wolf, Y. I., Roux, S., Camargo, A. P., Lee, B., Kazlauskas, D., Chen, I. M., Ivanova, N., Allen, L. Z., Paez-Espino, D., Bryant, D. A., Bhaya, D., Consortium, R. V. D., Krupovic, M., Dolja, V. V., Kyrpides, N. C., Koonin, E. V., and Gophna, U. 2022. A five-fold expansion of the global RNA virome reveals multiple new clades of RNA bacteriophages. bioRxiv:2022.2002.2015.480533.
Obbard, D. J. 2018. Expansion of the metazoan virosphere: Progress, pitfalls, and prospects. Curr. Opin. Virol. 31:17-23.
Olmos, A., Boonham, N., Candresse, T., Gentit, P., Giovani, B., Kutnjak, D., Liefting, L., Maree, H. J., Minafra, A., Moreira, A., Nakhla, M. K., Petter, F., Ravnikar, M., Rodoni, B., Roenhorst, J. W., Rott, M., Ruiz-García, A. B., Santala, J., Stancanelli, G., van der Vlugt, R., Varveri, C., Westenberg, M., Wetzel, T., Ziebell, H., and Massart, S. 2018. High-throughput sequencing technologies for plant pest diagnosis: Challenges and opportunities. EPPO Bull. 48:219-224.
Ondov, B. D., Bergman, N. H., and Phillippy, A. M. 2011. Interactive metagenomic visualization in a Web browser. BMC Bioinf. 12:385.
Palanga, E., Filloux, D., Martin, D. P., Fernandes, E., Gargani, D., Ferdinand, R., Zabre, J., Bouda, Z., Neya, J. B., Sawadogo, M., Traore, O., Peterschmitt, M., and Roumagnac, P. 2016. Metagenomic-based screening and molecular characterization of cowpea-infecting viruses in Burkina Faso. PLoS One 11:e0165188.
Palanga, E., Martin, D. P., Galzi, S., Zabre, J., Bouda, Z., Neya, J. B., Sawadogo, M., Traore, O., Peterschmitt, M., Roumagnac, P., and Filloux, D. 2017. Complete genome sequences of cowpea polerovirus 1 and cowpea polerovirus 2 infecting cowpea plants in Burkina Faso. Arch. Virol. 162:2149-2152.
Palanga, E., Tibiri, E. B., Bangratz, M., Filloux, D., Julian, C., Pinel-Galzi, A., Koala, M., Néya, J. B., Brugidou, C., Tiendrébéogo, F., Roumagnac, P., and Hébrard, E. 2021. Complete genome sequence of a novel marafivirus infecting pearl millet in Burkina Faso. Arch. Virol. 167:245-248.
Ren, J., Ahlgren, N. A., Lu, Y. Y., Fuhrman, J. A., and Sun, F. 2017. VirFinder: A novel k-mer based tool for identifying viral sequences from assembled metagenomic data. Microbiome 5:69.
Reynard, J.-S., Turco, S., Brodard, J., Kellenberger, I., Maclot, F., Schumpp, O., Gugerli, P., and Pooggin, M. M. 2021. Identification and molecular characterization of a novel hordeivirus associated with yellow mosaic disease of privet (Ligustrum vulgare) in Europe. Front. Microbiol. 12:723350.
Richards, R. S., Adams, I. P., Kreuze, J. F., De Souza, J., Cuellar, W., Dullemans, A. M., Van Der Vlugt, R. A., Glover, R., Hany, U., Dickinson, M., and Boonham, N. 2014. The complete genome sequences of two isolates of potato black ringspot virus and their relationship to other isolates and nepoviruses. Arch. Virol. 159:811-815.
Richet, C., Kraberger, S., Filloux, D., Bernardo, P., Harkins, G. W., Martin, D. P., Roumagnac, P., and Varsani, A. 2019. Novel circular DNA viruses associated with *Apiaceae* and *Poaceae* from South Africa and New Zealand. Arch. Virol. 164:237-242.
Roberts, J. M. K., Ireland, K. B., Tay, W. T., and Paini, D. 2018. Honey bee-assisted surveillance for early plant virus detection. Ann. Appl. Biol. 173: 285-293.
Roossinck, M. J. 2012. Plant virus metagenomics: Biodiversity and ecology. Annu. Rev. Genet. 46:359-369.
Roossinck, M. J., Martin, D. P., and Roumagnac, P. 2015. Plant virus metagenomics: Advances in virus discovery. Phytopathology 105:716-727.
Roossinck, M. J., Saha, P., Wiley, G., Quan, J., White, J., Lai, H., Chavarria, F., Shen, G., and Roe, B. 2010. Ecogenomics: Using massively parallel pyrosequencing to understand virus ecology. Mol. Ecol. 19:81-88.
Rosario, K., Capobianco, H., Ng, T. F., Breitbart, M., and Polston, J. E. 2014. RNA viral metagenome of whiteflies leads to the discovery and characterization of a whitefly-transmitted carlavirus in North America. PLoS One 9:e86748.
Rosario, K., Dayaram, A., Marinov, M., Ware, J., Kraberger, S., Stainton, D., Breitbart, M., and Varsani, A. 2012. Diverse circular ssDNA viruses discovered in dragonflies (Odonata: Epiprocta). J. Gen. Virol. 93:2668-2681.
Roumagnac, P., Granier, M., Bernardo, P., Deshoux, M., Ferdinand, R., Galzi, S., Fernandez, E., Julian, C., Abt, I., Filloux, D., Mesleard, F., Varsani, A., Blanc, S., Martin, D. P., and Peterschmitt, M. 2015. Alfalfa leaf curl virus: An aphid-transmitted geminivirus. J. Virol. 89:9683-9688.
Roumagnac, P., Lett, J. M., Fiallo-Olivé, E., Navas-Castillo, J., Zerbini, F. M., Martin, D. P., and Varsani, A. 2021. Establishment of five new genera in the family *Geminiviridae*: *Citlodavirus*, *Maldovirus*, *Mulcrilevirus*, *Opunvirus*, and *Topilevirus*. Arch Virol. 167:695-710.
Roux, S., Matthijnssens, J., and Dutilh, B. E. 2019. Metagenomics in virology. Reference module in Life Sciences: B978-970-912-809633-809638.820957-809636.
Ryckebusch, F., Sauvion, N., Granier, M., Roumagnac, P., and Peterschmitt, M. 2020. Alfalfa leaf curl virus is transmitted by *Aphis craccivora* in a highly specific circulative manner. Virology 546:98-108.
Sabanadzovic, S., Abou Ghanem-Sabanadzovic, N., and Valverde, R. A. 2010. A novel monopartite dsRNA virus from rhododendron. Arch. Virol. 155: 1859-1863.

Sabanadzovic, S., Valverde, R. A., Brown, J. K., Martin, R. R., and Tzanetakis, I. E. 2009. Southern tomato virus: The link between the families *Totiviridae* and *Partitiviridae*. Virus Res. 140:130-137.

Schloss, P. D., and Handelsman, J. 2005. Metagenomics for studying unculturable microorganisms: Cutting the Gordian knot. Genome Biol. 6:229.

Scussel, S., Candresse, T., Marais, A., Claverie, S., Hoareau, M., Azali, H. A., Verdin, E., Tepfer, M., Filloux, D., Fernandez, E., Roumagnac, P., Robène, I., Lefeuvre, P., Jourda, C., Roux-Cuvelier, M., and Lett, J.-M. 2019. High-throughput sequencing of complete genomes of ipomoviruses associated with an epidemic of cassava brown streak disease in the Comoros Archipelago. Arch. Virol. 164:2193-2196.

Sempéré, G., Pétel, A., Abbé, M., Lefeuvre, P., Roumagnac, P., Mahé, F., Baurens, G., and Filloux, D. 2021. metaXplor: An interactive viral and microbial metagenomic data manager. Gigascience 10:giab001.

Shi, M., Lin, X. D., Tian, J. H., Chen, L. J., Chen, X., Li, C. X., Qin, X. C., Li, J., Cao, J. P., Eden, J. S., Buchmann, J., Wang, W., Xu, J., Holmes, E. C., and Zhang, Y. Z. 2016. Redefining the invertebrate RNA virosphere. Nature 540:539-543.

Simmonds, P., Adams, M. J., Benko, M., Breitbart, M., Brister, J. R., Carstens, E. B., Davison, A. J., Delwart, E., Gorbalenya, A. E., Harrach, B., Hull, R., King, A. M., Koonin, E. V., Krupovic, M., Kuhn, J. H., Lefkowitz, E. J., Nibert, M. L., Orton, R., Roossinck, M. J., Sabanadzovic, S., Sullivan, M. B., Suttle, C. A., Tesh, R. B., van der Vlugt, R. A., Varsani, A., and Zerbini, F. M. 2017. Consensus statement: Virus taxonomy in the age of metagenomics. Nat. Rev. Microbiol. 15:161-168.

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

Sommers, P., Chatterjee, A., Varsani, A., and Trubl, G. 2021. Integrating viral metagenomics into an ecological framework. Annu. Rev. Virol. 8: 133-158.

Stobbe, A. H., and Roossinck, M. J. 2014. Plant virus metagenomics: What we know and why we need to know more. Front. Plant Sci. 5:150.

Susi, H., Filloux, D., Frilander, M. J., Roumagnac, P., and Laine, A. L. 2019. Diverse and variable virus communities in wild plant populations revealed by metagenomic tools. PeerJ 7:e6140.

Susi, H., Laine, A. L., Filloux, D., Kraberger, S., Farkas, K., Bernardo, P., Frilander, M. J., Martin, D. P., Varsani, A., and Roumagnac, P. 2017. Genome sequences of a capulavirus infecting *Plantago lanceolata* in the Aland archipelago of Finland. Arch. Virol. 162:2041-2045.

Varsani, A., and Krupovic, M. 2021. Family Genomoviridae: 2021 taxonomy update. Arch. Virol. 166:2911-2926.

Victoria, J. G., Kapoor, A., Dupuis, K., Schnurr, D. P., and Delwart, E. L. 2008. Rapid identification of known and new RNA viruses from animal tissues. PLoS Pathog. 4:e1000163.

Walker, P. J., Siddell, S. G., Lefkowitz, E. J., Mushegian, A. R., Adriaenssens, E. M., Alfenas-Zerbini, P., Davison, A. J., Dempsey, D. M., Dutilh, B. E., García, M. L., Harrach, B., Harrison, R. L., Hendrickson, R. C., Junglen, S., Knowles, N. J., Krupovic, M., Kuhn, J. H., Lambert, A. J., Łobocka, M., Nibert, M. L., Oksanen, H. M., Orton, R. J., Robertson, D. L., Rubino, L., Sabanadzovic, S., Simmonds, P., Smith, D. B., Suzuki, N., Van Dooerslaer, K., Vandamme, A.-M., Varsani, A., and Zerbini, F. M. 2021. Changes to virus taxonomy and to the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2021). Arch. Virol. 166:2633-2648.

Whattam, M., Dinsdale, A., and Elliott, C. E. 2021. Evolution of plant virus diagnostics used in Australian post entry quarantine. Plants 10:1430.

Wren, J. D., Roossinck, M. J., Nelson, R. S., Scheets, K., Palmer, M. W., and Melcher, U. 2006. Plant virus biodiversity and ecology. PLoS Biol. 4:e80.

Wu, S., and Manber, U. 1992. A fast approximate pattern-matching tool. Pages 153-162 in: Usenix Winter 1992 Technical Conference, San Francisco.

Zhang, Y. Z., Shi, M., and Holmes, E. C. 2018. Using metagenomics to characterize an expanding virosphere. Cell 172:1168-1172.