

Metagenomic Analyses of Nepoviruses: Diversity, Evolution and Identification of a Hitherto Undescribed Putative Region for Host Range in Subgroup a Species

jean-michel hily (✉ jean-michel.hily@vignevin.com)

IFV: Institut Francais de la Vigne et du Vin <https://orcid.org/0000-0001-9141-9701>

Nils Poulicard

IRD: Institut de recherche pour le developpement

Julie Kubina

Institut National de Recherche pour l'Agriculture l'Alimentation et l'Environnement Centre Grand-Est
Nancy

Jean-sebastien Reynard

ACW Standort Changins: Agroscope Standort Changins

Anne-Sophie Spilmont

IFV: Institut Francais de la Vigne et du Vin

Marc Fuchs

Cornell University New York State Agricultural Experiment Station: Cornell AgriTech

Olivier Lemaire

INRAE Centre Grand-Est Colmar: Institut National de Recherche pour l'Agriculture l'Alimentation et
l'Environnement Centre Grand-Est Nancy

Emmanuelle Vigne

INRAE Centre Grand-Est Colmar: Institut National de Recherche pour l'Agriculture l'Alimentation et
l'Environnement Centre Grand-Est Nancy

Original Article

Keywords: Metagenomic Analyses of Nepoviruses, Hitherto Undescribed Putative,

Posted Date: February 8th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-171249/v1>

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

[Read Full License](#)

Version of Record: A version of this preprint was published at Archives of Virology on August 9th, 2021.
See the published version at <https://doi.org/10.1007/s00705-021-05111-0>.

Metagenomic analyses of nepoviruses: diversity, evolution and identification of a hitherto undescribed putative region for host range in subgroup A species

Hily JM^{1,2*}, Poulicard N³, Kubina J², Reynard JS⁴, Spilmont AS¹, Fuchs M⁵, Lemaire O² and Vigne E^{2*}.

¹ IFV, Le Grau-Du-Roi, France

² Université de Strasbourg, INRAE, SVQV UMR-A 1131, F-68000 Colmar, France

³ PHIM, Université Montpellier, IRD, INRAE, Cirad, SupAgro, Montpellier, France

⁴ Institute for Plant Production Science, Agroscope, 1260 Nyon, Switzerland

⁵ Cornell University, Geneva, NY, USA

* Corresponding authors: jean-michel.hily@vignevin.com; emmanuelle.vigne@inrae.fr

ORCID: Hily 0000-0001-9141-9701; Poulicard 000-0003-3969-2131; Vigne 0000-0002-3287-0786.

ABSTRACT

Datamining and metagenomic analyses of 277 open reading frame sequences of bipartite RNA viruses and variants in the genus *Nepovirus* documented how delicate it can be to unequivocally identify species, in particular subgroup A and C species, based on some of the currently adopted taxonomic demarcation criteria. It suggests a possible need for their amendment to accommodate pangenome information. In addition, we revealed a host-dependent structure of arabis mosaic virus (ArMV) populations at a cladistic level and confirmed a phylogeographic structure of grapevine fanleaf virus (GFLV) populations. We also identified new putative recombinant events for species of subgroups A, B and C. The evolutionary specificity of some capsid regions of ArMV and GFLV that were previously described and biologically validated as vector determinant was circumscribed *in silico*. Furthermore, a C-terminal segment of the RNA-dependent RNA polymerase of subgroup A species was predicted as a putative host range determinant based on statistically supported higher π values for GFLV and ArMV isolates infecting *Vitis* spp. compared to non-*Vitis* infecting ArMV isolates. This study illustrated how sequence information obtained *via* high throughput sequencing can increase our understanding of mechanisms that modulate virus diversity and evolution and create new opportunities for advancing studies on the biology of economically important plant viruses.

INTRODUCTION

New viral sequences are uncovered at an unprecedented rate with the advent of high throughput sequencing (HTS). The recovery of numerous complete or almost complete viral

genome sequences from different ecosystems (*i.e.*, environmental, human, veterinary, plant), lead to metagenomic analyses that assist the description of unforeseen virus genomes and their diversity. This wealth of information is creating the possibility to use the pangenome for virus taxonomy [17] and increase our understanding of the mechanisms that modulate virus diversity, evolution, vector and host specificity and epidemiology [36]. However, new challenges arise, for instance, for virus classification. Taxonomy traditionally relies not only on the genetic relationship among sequences of a few virus coding regions, primarily the replicase and/or coat protein coding domains, but also on biological properties such as vector species and host range, among other features [46]. This type of biological information is critical for current plant virus taxonomic classification but is generally lacking when only metagenomic data are available.

Nepoviruses are plant picorna-like viruses of the subfamily *Comovirinae* in the family *Secoviridae* [44]. Their transmission occurs in a non-persistent and non-circulative manner by ectoparasitic nematodes from the genera *Xiphinema*, *Longidorus* and *Paralongidorus* [42]. Long distance dissemination of nepoviruses happens with the exchange of uncontrolled propagation material and the use of infected cuttings and budwood for grafting. Seed and pollen transmission have been documented for some but not all nepoviruses and transmission by mites has been observed in rare cases. The genus *Nepovirus* includes 40 species that are widely distributed in temperate regions (https://talk.ictvonline.org/ictv-reports/ictv_online_report/positive-sense-rna-viruses/picornavirales/w/secoviridae/591/genus-nepovirus) [22]. Most nepoviruses have a broad natural host range, including annual herbaceous species and perennial woody species, and cause significant crop losses worldwide [11].

The genome of nepoviruses is composed of two single-stranded, positive sense RNAs (RNA1 and RNA2). Both genomic RNAs are necessary for infection *in planta*. They encode a large polyprotein, P1 for RNA1 and P2 for RNA2, which is cleaved by the viral proteinase into functional proteins [11]. P1 encodes proteins that are necessary for replication, including a helicase with a nucleotide triphosphate-binding domain, a proteinase (Pro) and an RNA-dependent RNA polymerase (Pol). Depending on the viral species, one (1A) or two (X1 and X2) proteins are located upstream of the helicase domain. The function of these proteins is not fully elucidated yet. P2 contains the coat protein (CP) for which the assembly of multiple units form icosahedral virions of 26-30 nm in diameter. The cell-to-cell movement (MP) domain is located immediately upstream of the CP domain. Depending on the nepovirus species, either one (2A^{HP}, which is required for the replication of RNA2) or two (X3 and X4 of unknown functions) proteins are located upstream of the MP [13]. Three subgroups of nepoviruses are recognized based on RNA2 properties, including its organization and size, phylogenetic relationships of the CP coding region, and cleavage sites recognized by the viral proteinase [11]. The three nepovirus subgroups are named A, B and C.

One of the most important viral diseases of grapevines is infectious degeneration. This disease is caused by 15 *Nepovirus* species [6, 41]. Most grapevine-infecting nepoviruses are generally restricted to some regions of the world. For example, arabis mosaic virus (ArMV) is

limited to European vineyards while tobacco ringspot virus (TRSV), tomato ringspot virus (ToRSV), peach rosette mosaic virus (PRMV) and blueberry leaf mottle virus are present in American vineyards. In contrast, grapevine fanleaf virus (GFLV) is present in most vineyards worldwide.

Genetic diversity of nepoviruses has been extensively characterized with information primarily collected from RT-PCR-based studies combined with Sanger sequencing, generally in the part of the genome that contains the coat protein coding region [14, 50]. Similarly, diversity studies and phylogenetic analyses have been reported for members of the family *Secoviridae*, including nepoviruses [43, 49]. However, several new nepovirus species have been recently characterized and the number of complete genome sequences of many nepovirus isolates has exponentially increased in the past five years [1, 2, 4, 12, 15, 18, 23-25, 39, 48, 53, 54, 56]. In this study, we built on these latest advancements in nepoviruses research and carried out metagenomic analyses. We focused on RNA1 and RNA2 coding sequences to gain new insights into viral diversity and evolution, and to predict a hitherto undescribed conserved region putatively determining the host range of two subgroup A nepoviruses.

RESULTS AND DISCUSSION

Phylogenetic relationships among nepoviruses

Only complete or near complete (*i.e.*, covering ORF1 or ORF2) sequences of nepoviruses RNA1 and RNA2 were considered in this study. All sequences were retrieved from NCBI as of January 2020, our own curated nepoviruses sequences repository and a selection of Sequence Read Archive (SRA) datasets from GenBank. Datamining was performed to increase the number of sequences for ArMV, GFLV and mulberry mosaic leafroll-associated virus (MMLRaV), a novel *Nepovirus* species [31], as previously described [19]. These datamining efforts and Sanger or Illumina sequencing resulted in 46 new sequences (24 for RNA1 and 22 for RNA2) of ArMV, GFLV and MMLRaV. These new sequences were deposited in GenBank ([Table S1](#), *** Submission has been done. However, no GenBank ID yet.***). In total, both genomic RNA sequences were recovered from 29 *Nepovirus* species except from olive latent ringspot virus for which only a single RNA2 sequence but no RNA1 sequence is available ([Table 1](#)). Two species (GFLV and ArMV) made up the majority of sequences analyzed in this study, while most species were represented by a single or a few sequences of either genomic RNAs ([Table 1](#)). Novel nepoviruses used in this study included MMLRaV [31], caraway yellows virus [12], potato virus B [4] and red clover nepovirus A [25]. A few new virus species and new isolates belonging to the genus *Nepovirus* have been identified since we last consulted NCBI (April 2020). The corresponding sequences were not included in this study ([Table S10](#)). In addition, a few viral species described in the literature as potential new nepoviral species such as Hobart nepovirus 3 [40] or Zhuye pepper nepovirus 1 [3] were not taken into account in this study because sequences were partial or discrepancies were observed between datasets available in NCBI and publications. Furthermore, only a single sequence was chosen from a group of sequences displaying nucleic acid identity higher than 99%,

unless isolates were obtained from different hosts and/or different countries. The complete list of 110 ORF1 (Table S2) and 167 ORF2 (Table S3) nepovirus sequences selected for this study is provided.

Genetic differentiation analyses of nucleotide sequences confirmed the classification of nepoviruses into three subgroups with higher inter-subgroup than intra-subgroup mean distance values (Table 2). Subgroup B sequences displayed the lowest maximum pairwise distance values, well below inter-subgroup mean distance values, suggesting a well-defined group of virus isolates (Table 2). The inter-subgroup mean distance value was lower than the maximum intra-subgroup mean distance value for subgroup A and C sequences, revealing a greater variability and less well-defined groups of virus isolates (Table 2). After performing an alignment of ORF1 and ORF2 sequences, phylogenetic trees were constructed with maximum-likelihood method using the best-fit model (GTR+G+I) (Figure 1). Interestingly, species belonging to each subgroup separated better in a tree considering ORF1 rather than ORF2 sequences (Figure 1 and Figure S1). Indeed, ORF1 nucleotide sequences of virus isolates of subgroups A, B and C clustered in separate clades in a tanglegram (Figure 1) or unrooted cladogram (Figure S1). Nucleotide sequences of nepovirus isolates of subgroup B were also well-defined when using ORF2 sequences, however, subgroup A and C ORF2 sequences were scattered in different clades in a tanglegram (Figure 1) or unrooted cladogram (Figure S1). These results suggested that criteria used for the classification of nepoviruses into distinct subgroups are more robust based on ORF1 than ORF2 nucleotide sequence identity. This finding could be considered by the International Committee on Taxonomy of Viruses (ICTV) working group on *Secoviridae* to eventually define new demarcation criteria for nepoviruses when using pangenome information.

New challenges for species identification within the genus *Nepovirus*

Species demarcation criteria for nepoviruses are defined by the ICTV (<https://talk.ictvonline.org/ictv-reports/ictv-online-report/positive-sense-rna-viruses/picornavirales/w/secoviridae>). These include coat protein (CP) amino acid (aa) sequence identity less than 75% and conserved Protease-Polymerase (Pro-Pol) region aa sequence identity less than 80%, among other criteria. We assessed whether these two major demarcation criteria are applicable to the corresponding complete ORF1 and ORF2 aa sequences. Some discrepancies with regard to the intra-species aa identity falling outside the species demarcation were obtained for PRMV ORF1 sequences (78.99%) and ORF2 sequences of cherry leaf roll virus (CLRV), ToRSV, cycas necrotic stunt virus and ArMV (below 74.05%) (Table S4). These results revealed that analyzing complete ORF sequences may be problematic for the identification of new virus species and new genetic variants of existing virus species if the current demarcation criteria pertaining to partial genome sequence information were to be applied. They further suggested that demarcation criteria for species in the genus *Nepovirus* could be amended to accommodate

pangenome information. In addition, the ORF2 sequence of ArMV isolate Butterbur was a clear outlier among ArMV isolates with distinct identities at both the nucleic acid (Figure 2) and amino acid (70.25% Table S4) levels. According to the original report [21], the pathological and serological features of ArMV-Butterbur are unique and its coat protein is 504 aa long (as for all GFLV CPs), while all other ArMV CPs are 505 aa long. These features underscore the need for additional work to ascertain the taxonomic position of ArMV-Butterbur and its recognition as an isolate of the ArMV species.

Similarly, the ORF1 amino acid sequence identity between isolates of distinct species was higher than 80%, for example for beet ringspot virus (BRSV) and tomato black ring virus (TBRV), as well as for BRSV and artichoke Italian latent virus (AILV) (Table S4). The high identity percentage could also explain the high number of inter-species recombination events identified between these particular species (see the dedicated section below). However, inter-species diversity was below the species demarcation level (< 75%) for ORF2 sequences, unquestionably defining BRSV, TBRV and AILV as different species (Table S4). One particular case of interest is grapevine deformation virus (GDefV) [20], a subgroup A nepovirus. GDefV ORF2 aa sequences display 73 and 71% homologies with GFLV and ArMV, respectively [16], and GDefV ORF1 aa sequences have higher homology with GFLV (86-89%) than ArMV (73-74%) sequences [8]. According to the species demarcation criteria for nepoviruses, GDefV could be classified as a highly divergent sequence of GFLV when focusing on ORF1 sequences but as a new species based on ORF2 sequences.

Identification of recombination events within and between nepovirus species

Putative intra-species recombination events have been extensively reported for nepoviruses, mostly in the GFLV RNA2-encoded movement protein (MP) and CP domains [33, 34, 37, 47, 50-52]. Recombination events have also been described for ToRSV [54] and grapevine chrome mosaic virus (GCMV) [5]. In addition, many inter-species recombination events have been detected, mostly between ArMV and GFLV [9, 33, 34, 52] but also between GCMV and TBRV [5, 28]. With the use of HTS and the recovery of complete virus genomes, recombination events can be detected all along the two genomic RNAs [18]. Here, we used the same corpus of nepovirus sequences and searched for evidence of recombination events using the RDP4 program. Recombination events were only considered when detected with five or more methods with $P < 10^{-3}$ (Table 3, Tables S5 and S6).

Intra-species recombination events were identified in ORF1 and ORF2 sequences, mostly of subgroup A species (Table 3, Table S5 and Figure S2). Almost twice as many putative recombination events were detected in ORF2 compared to ORF1 sequences (Table 3). Both remarks most definitely reflect the total number of sequences being recovered and used in this study. Many recombination events were detected in GFLV and ArMV sequences with some hotspots, *i.e.* more than one recombinant per site (Table S5 and Figure S2). In addition,

recombination events were also identified for the first time for AILV, CLRV and raspberry ringspot virus.

All inter-species recombination events detected in this study strictly involved species from the same subgroup (Table S6). Surprisingly, the number of inter-species recombination events was higher than intra-species recombination events within ORF2 sequences (Table 3). For example, 31 inter-species and two intra-species recombination events were detected for subgroup B ORF2 sequences (Table 3 and Figure S2). It should also be noted that all ORF1 recombination events detected for subgroup B involved different species (Table S6). In contrast, all subgroup A recombination events detected involved only ArMV, GFLV and GDefV (Figure 1 and Figure S1), emphasizing again their kinship. Recombination may have been facilitated for these three species because they have the potential to exist or co-exist in grapevine, a common host, for a long time, thus increasing their potential encounter in a single host cell.

Genetic diversity and population differentiation of ArMV from mono- and dicotyledonous plants

ArMV is a 'generalist' with a very broad natural host range, including winter barley, narcissus, *Ligustrum vulgare*, weeds, hops, berries and grapevines, among other species. Our datamining efforts resulted in the retrieval of 17 complete ORF1 sequences from seven monocotyledonous plants and 10 dicotyledonous plants, as well as 21 complete ORF2 sequences from eight monocotyledonous plants and 13 dicotyledonous plants (Tables S1, S2 and S8).

The overall genetic diversity (π) of ArMV ORF1 and ORF2 was of 0.166 ± 0.003 and 0.133 ± 0.003 , respectively (Table 4). As previously observed [55], the coding region 2A^{HP} is the most divergent genomic region, showing the highest diversity at the extreme 5' end of ORF2 sequences (Figure 2), mostly due to size differences among isolates. A comparative analysis of ArMV sequences obtained from mono- and dicotyledonous plants revealed a significantly higher diversity in sequences from isolates infecting dicotyledonous plants compared to isolates infecting monocotyledonous plants (0.170 ± 0.003 vs 0.109 ± 0.003 and 0.145 ± 0.003 vs 0.093 ± 0.003 , for ORF1 and ORF2 sequences respectively, Table 4). When looking at the evolution pattern (Tajima's D) of ORF1 sequences (Figure 2), values were negative but close to 0 ($D_T = -0.344$; $P > 0.1$), suggesting that the population of ArMV is evolving as per mutation-drift equilibrium with no specific region under selection. On the other hand, two distinct regions of ORF2 sequences were under selection with an overall D_T of -0.994 ($P > 0.1$) (Figure 2). One of these two regions is covering an aa stretch between two proline-rich segments of the central coding region of the 2A^{HP} domain. The other region is a specific segment of the 2C^{CP} coding region that overlaps the previously defined R4 region, which is involved in specific ArMV transmission by the nematode vector, *Xiphinema diversicaudatum* [45].

In a previous study [55], ArMV isolates were separated by the size and aa sequence identity of protein 2A^{HP} into four groups (I to IV). Here, we recovered 43 2A^{HP} nucleotide sequences from GenBank (Table S8) and confirmed the existence of three major clades

corresponding to groups II, III and IV (Figure S3). Group I was composed of a single sequence located within the group II clade. Sequences belonging to each group are genetically different with a high fixation index ($F_{ST} \geq 0.530$) and strong statistical support ($P \leq 10^{-5}$) (Table 5). However, the size of the 2A^{HP} domain was not linked to the plant host with ArMV isolates from grapevine belonging to all four groups. A comparative analysis of 2A coding sequences from mono- and dicotyledonous plants documented a statistically supported genetic differentiation (F_{ST}) (Table 5). Genetic differentiations according to mono- and dicotyledonous plants were also observed when looking at any other RNA1 or RNA2 coding region sequences or the complete ORF1 and ORF2 sequences (Table 6, Tables S7-S8 and Figure S3). Distinct F_{ST} between mono- and dicotyledonous plants were also found at lower cladistical levels, strongly suggesting a likely genetic bias based on the plant host (Table S7). Interestingly, similar results have been reported for CLRV, another generalist virus within the genus *Nepovirus*, for which a host species-dependent population structure was documented using only a short 375 pb sequence corresponding to the extreme 3' part of the 3' non coding region [38].

Genetic diversity and population differentiation of GFLV from different geographic origins

GFLV infects primarily *Vitis* spp., making the virus very specialized to this woody plant. The overall nucleotide diversity for GFLV ORF1 ($\pi = 0.127 \pm 0.002$) and ORF2 ($\pi = 0.130 \pm 0.005$) sequences was very similar (Table 4). Plotting π along ORF1 sequences (Figure 3) showed a highly divergent region at the 3' end of the 1E^{Pol} domain. This result is consistent with other analyses of this particular aa stretch of P1 that was predicted to be in a helix at the aa level [18, 35]. Another highly polymorphic region was detected at the extreme 5' end of ORF2 sequences (Figure 3), which corresponds to a region where intra- and inter-species recombination events have been predicted (see above section and [52]). On the other hand, similar to ArMV, a significant drop in nucleotide diversity is observed within a segment of 2A^{HP} sequences located between two highly conserved proline-rich regions. The evolution of this particular domain of ORF2 sequences is not neutral with statistical D_T values well below 0 (Figure 3), underlining a conservative selection with regard to the remainder of ORF2 sequences. Similarly to ArMV, the same trend was observed for the R4 region of the 2C^{CP} domain [45]. Interestingly, these two regions, which display the lowest D_T values suggesting a recent selective sweep, are mostly located in sequences recovered from grapevines from the New World (Figure S4). Regarding the evolution pattern of GFLV ORF1 sequences, values were negative but very close to 0 ($D_T = -0.758$) with two sites under selection ($P > 0.1$): the first site is located at the extreme 5' end and the second site is positioned within the 1B^{Hel} domain (Figure 3). When looking at the evolutionary pattern of P1 and P2 (dN-dS), most codons were under negative or neutral pressure (data not shown), as previously described [49].

No major differences were observed when separating GFLV sequences by geographic region (France vs rest of the world or old vs new world), with very similar π and D_T values (Table 4 and Figure S4). However, a genetic structuration between geographic regions was observed although F_{ST} values were extremely low (Table S9) with differences in the evolution pattern of

GFLV isolates from different part of the world. This was even more noticeable when focusing on ORF2 sequences and separating sequences by continent [*i.e.*, Europe, Americas (combining North and South America) and Asia (far-East, Turkey and Russia)]. All F_{ST} were statistically supported ($P < 0.001$). However, the disparity in F_{ST} values indicated that European and American GFLV variants were closely related to each other compared to the Asian variants. This observation was confirmed when grouping sequences into seven countries or restricted regions of the world (France, Slovenia, Italy, USA, Chile, far-East and Switzerland). Some F_{ST} values were very high, underlying a strong genetic structuration among regions around the world, as confirmed when comparing sequences from the far-East (Iran and China) with other regions of the world (Table S9). This genetic differentiation according to far-East GFLV populations was previously described using GFLV 2B^{MP} sequences [47]. Altogether, these observations suggest a specific geographic evolution and genetic structuration of the virus.

Common characteristics and major differences between grapevine-infecting ArMV and GFLV isolates.

ArMV and GFLV are very closely related but different species (Figure 1). They share many characteristics such as hosts (*i.e.*, grapevine), closely related vectors (*Xiphinema* spp.), similar symptomatology and many natural inter-species recombinants. While genetically different (Table S4, Figures 4 and 5), similar patterns in their respective genetic diversity were observed along ORF1 sequences, especially when decomposing grapevine-infecting ArMV isolates from non-*Vitis*-infecting ArMV isolates. As detailed above, one of the hallmarks of GFLV is a higher π value at the C-terminus of 1E^{Pol}. A higher π at the C-terminal end of 1E^{Pol} was also clearly identified in *Vitis*-infecting but not in non-*Vitis*-infecting ArMV isolates (Figure 4). Such specific increased genetic diversity in *Vitis*-infecting ArMV and GFLV isolates was not due to the overlap of a hidden ORF (Figure S6), as described for sobemoviruses [30]. This diversity was also observed at the amino acid level with a percent identity higher than 80.41% in the case of non-*Vitis*-infecting ArMV isolates, but as low as 65.54% and 56.08% for *Vitis*-infecting ArMV and GFLV isolates, respectively (Figure 4). Such highly divergence was not found when specifically looking at the first 148 aa of the 1E^{Pol} domain with an identity above 82%. While highly divergent between ArMV and GFLV (Figure S7), the 1E^{Pol} C-terminus displays only two aa conserved between *Vitis*-infecting ArMV and GFLV isolates at position 683 and 746 (Figure S8). Could these two residues be implicated in host adaptation mechanisms? More work is needed to address this hypothesis.

Many characteristics are common between ORF2 sequences of ArMV and GFLV (Figure 5). For example, a higher genetic diversity is detected at the 5' end of the 2A^{HP} domain, partly as a result of indels. However, major differences between the two species were found when focusing specifically on the 2C^{CP} domain with a clearly different genetic diversity in the R4-R5 region between ArMV and GFLV (Figure 5 and Figure S5). This region is important for vector transmission [45]. When looking at the evolution pattern, most of the ORF2 sequences seem to be evolving randomly while two regions display non-random evolution pattern. One of these two regions

corresponds to the 2A^{HP} domain and the other to the R4-R5 region within the 2C^{CP} domain, both showing strong constraints.

CONCLUSION

Datamining and metagenomic efforts of complete ORF sequences unveiled new insights into the diversity and evaluation of virus species from the genus *Nepovirus* with a special emphasis on GFLV and ArMV, the two most important viruses involved in degeneration disease of grapevine in France. Our results confirmed a probable phylogeographic structure of GFLV populations and revealed a host-dependent structure of ArMV populations at a cladistic level. The C-terminus of the RNA-dependent RNA polymerase of GFLV and ArMV is predicted as a potential host range determinant. More work is needed to biologically test this hypothesis. Furthermore, some of the current species demarcation criteria that are applied to limited genomic regions may not be validated for all nepoviruses at the ORF sequence level. This suggested the need to adapt some of the taxonomic criteria to pangenome information. Nonetheless, with an ever-increasing amount of sequence data obtained through HTS, there are new opportunities for studying *Nepovirus* biology, characterizing nepoviral communities in plants, improving *Nepovirus* taxonomy and confronting these new pangenomic and populational information for developing resistances strategies (*i.e.*, cross-protection, resistance alleles characterization, new plant biotechnological technologies) and challenging them for a better durability of protection.

MATERIALS & METHODS

Sequence analyses, genetic diversity and recombination detection

The complete to near complete nepovirus ORF1 and ORF2 sequences were retrieved from NCBI as of April 2020, our own curated nepovirus sequences repository obtained by analyses of high throughput or Sanger sequencing datasets, and selection of Sequence Research Archive datasets from GenBank [19]. In total, 110 ORF1 sequences and 167 ORF2 sequences have been used in this study (Table S2 and S3). In addition, sequences of specific domains we retrieved from NCBI (see Table S8).

Multiple sequence alignments by codon and maximum likelihood (ML)-based phylogenetic trees were prepared using MUSCLE [7] implemented in MEGA7 and MEGAX software [26, 27], excluding the viral untranslated regions (UTRs). The best ML-fitted model for each sequence alignment was used, and nodes in phylogenetic trees were validated by bootstrap analyses (100 replicates). For visualization effects, FigTree v. 1.3.1 was used (<http://tree.bio.ed.ac.uk/>). The diversity index (π), which is the average number of nucleotide substitutions per site between any two sequences in a multi-sequence alignment, and the variation of π along genome sequences were evaluated by sliding window analyses (length: 80, step size: 20) using DnaSP v.6.12.03 [29] and MegaX.

A search for potential recombination signals was performed using all seven algorithms implemented in the RDP program v4.97 [32]. The default settings were used for each algorithm and only recombination events detected by five or more methods were considered.

Differences in nucleotide diversity of viral populations defined using different modalities were tested by analysis of molecular variance (AMOVA), as implemented in Arlequin v. 5.3.1.2 [10]. AMOVA calculates the F_{ST} index explaining the between-groups fraction of total genetic diversity. The significance of these differences was evaluated by performing 1000 sequence permutations.

Tajima's D (D_T) and sliding window analyses were conducted using DnaSP v. 6.12.03 [29] in order to distinguish the viral populations evolving randomly (per mutation-drift equilibrium; $D_T = 0$) from those evolving under a nonrandom process ($D_T > 0$: balancing selection, sudden population contraction; $D_T < 0$: recent selective sweep, population expansion after a recent bottleneck).

LEGENDS

Table 1. List of nepoviral species used in this study. Species from each subgroup (A in red, B in blue and C in green) are indicated by a bold capital letter when officially ratified by ICTV or in lowercase when ICTV ratification is pending. The number of ORF1 and ORF2 sequences used for each species is provided. The length of both ORFs is indicated for each species. * indicates a shorter sequence for an isolate of cycas necrotic stunt virus (reference isolate) and red clover nepovirus A. na: not applicable.

Table 2. Genetic distance within and between subgroup species of the genus *Nepovirus*. Mean nucleotide distance is indicated in bold and standard error in italic. The maximum value of pairwise distance within subgroups is provided. N corresponds to the number of sequences used for calculation.

Table 3. Number of putative intra- and inter-species recombination events detected by RDP4 for species of the three subgroups in the genus *Nepovirus*. Detailed informations on the genomic location of recombination events, major and minor parents and P -values is provided in Tables S5 and S6.

Table 4. Genetic diversity for both ORFs of arabis mosaic virus (ArMV) and grapevine fanleaf virus (GFLV) isolates. Overall diversity index (π) \pm standard error (SE) and Tajima's D (D_T) with associated P values based on N (number of sequences per group) are provided. Sequence populations were grouped according to the plant type or the geographic origin of the isolates (monocot: monocotyledonous; dicot: dicotyledonous; *Vitis*; non-*Vitis*; Old: old world (France, Hungary, Italy, Switzerland); New: new world (Canada, USA, China, South Africa); FR: France; RoTW: rest of the world other than FR; EU: Europe; Am: Americas; As: Asia; IT: Italy; SL: Slovenia;

CA: Canada; CH: Switzerland; CL: Chile and FE: far East. The geographic origin of ArMV and GFLV isolates is specified for each sequence in Tables S2 and S3.

Table 5. Genetic differentiation of arabis mosaic virus (ArMV) populations for the complete ORFs and the different coding regions. Fixation index (F_{ST}) with associated P-value (p -val are significant when < 0.05) and the number of sequences (N) are indicated. Sequences were grouped by either the plant type (monocotyledonous versus dicotyledonous) or the size of the 2A^{HP} coding region (group II, III and IV).

Figure 1. Tanglegram of maximum likelihood phylogenetic trees inferred from 110 ORF1 and 167 ORF2 nucleotide sequences of nepoviruses. Colors correspond to species affiliation to one of the three nepovirus subgroups: red for subgroup A sequences; blue for subgroup B sequences; and green for subgroup C sequences. Clades with several sequences from the same species are collapsed. Numbers at nodes indicate bootstrap values based on 100 replicates. The scale bar corresponds to the number of substitutions per site.

Figure 2. Phylogenetic and diversity analyses of arabis mosaic virus (ArMV) isolates from different plants using a corpus of 17 ORF1 (left panel) and 21 ORF2 (right panel) nucleotide sequences. Colors correspond to hosts with red for monocotyledonous plants, green for dicotyledonous plants and black for all the plants. Maximum likelihood phylogenetic trees are shown. Numbers at each node indicate bootstrap values based on 100 replicates and scale bars show genetic distance. Graphics represent π (substitution/sites) and Tajima's D (D_T) for evolution along the ORF1 and ORF2 sequences. Colored bars with # and * correspond to statistically validated regions (P-values at 0.05 and 0.001, respectively).

Figure 3. Genetic diversity analyses of grapevine fanleaf virus (GFLV) isolates using a corpus of 40 ORF1 (left panel) and 80 ORF2 (right panel) nucleotide sequences. Graphics represent π (substitution/sites) and Tajima's D (D_T) for evolution along the ORF1 and ORF2 sequences. Bars and # and * correspond to statistically validated regions (P-values at 0.05 and 0.001, respectively).

Figure 4. Phylogenetic and diversity analyses of grapevine fanleaf virus (GFLV), arabis mosaic virus (ArMV) and grapevine deformation virus (GDefV) using a corpus of 58 ORF1 nucleotide sequences. GFLV isolates are in blue, GDefV isolates in peach, ArMV isolates from grapevines (Vitis-ArMV) in green and ArMV isolates from other plants (non-vitis ArMV) in red. Maximum likelihood phylogenetic trees are shown. Numbers at each node indicate bootstrap values based on 100 replicates and the scale bar shows genetic distance. Graphics represent π (substitution/sites) and Tajima's D (D_T) for evolution along the ORF1 sequence. Percentages correspond to minimum aa identity for both framed regions within the polymerase domain. The colored bar with # corresponds to a statistically validated region at 0.05.

Figure 5. Phylogenetic and genetic diversity analyses of grapevine fanleaf virus (GFLV), arabis mosaic virus (ArMV) and grapevine deformation virus (GDefV) using a corpus of 102 ORF2

nucleotide sequences. GFLV isolates are in blue, GDefV in peach, ArMV isolates from grapevines (Vitis-ArMV) in green and ArMV isolates from other plants (non-vitis ArMV) in red. Maximum likelihood phylogenetic trees are shown. Numbers at each node indicate bootstrap values based on 100 replicates and the scale bar shows genetic distance. Graphics represent π (substitution/sites) and Tajima's D (D_T) for evolution along the ORF2 sequence. The boxed area corresponds to the R4-R5 region of the 2C^{CP} domain. Colored bars with # and * correspond to statistically validated regions (P-values at 0.05 and 0.001, respectively).

Supp Table 1. List of the new nepoviral sequences obtained by datamining or from our own curated sequences repository that were released in this study. Virus name, acronym, specific ID, country of origin, sampling year and sequence release date and plant host is indicated for RNA1 or RNA2 sequences.

Supp Table 2. Complete list of the 110 nepoviral ORF1 sequences used in this study. Accession number, virus name, acronym, specific ID (NCBI), country of origin, sampling year (or paper publication date), NCBI release date, plant host species are provided.

Supp Table 3. Complete list of the 167 nepoviral ORF2 sequences used in this study. Accession number, virus name, acronym, specific ID (NCBI), country of origin, sampling year (or paper release date), NCBI release date, plant host species are provided.

Supp Table 4. Amino acids identity within and between species for which the demarcation is ambiguous under the criteria of less than 80% for ORF1 and 75% for ORF2. Percentages between species higher or lower than the criteria are highlighted in grey and bold respectively.

Supp Table 5. Intra-species recombination events detected on nepoviral ORF1 and ORF2 sequences using Recombination Detection Program package (RDP v.4.97). The species name is provided with the number of recombinant events, recombinant ID (the first one listed by RDP4 if multiple events were detected), major and minor parents with respective similarity percentages, recombination event location without gaps, number of programs identifying the event, names of the programs (R: RDP, B: Bootscan, G: GeneConv, M: MaxChi, C: Chimaera, S: Siscan and 3: 3Seq) and associated *P*-value.

Supp Table 6. Inter-species recombination events detected on nepoviral ORF1 and ORF2 sequences using Recombination Detection Program package (RDP v.4.97). The species name is provided with the number of recombination events in parenthesis for that particular location, type, recombinant ID (the first one listed by RDP4 if multiple events were detected), major and minor parents with respective similarity percentages, recombination event location without gaps, number of programs identifying the event, names of the programs (R: RDP, B: Bootscan, G: GeneConv, M: MaxChi, C: Chimaera, S: Siscan and 3: 3Seq) and associated *P*-value.

Supp Table 7. Differentiation measurements differentiation (fixation index, F_{ST}) and associated statistics (*P*-value) of arabis mosaic virus (ArMV) populations for ORF1 and ORF2 sequences and different coding regions. Isolates were categorized according to the type of plants

infected (monocotyledonous, dicotyledonous and lower cladistical levels). The number of sequences (N) for each category is indicated. Only statistically validated F_{ST} are provided (p-value < 0.05).

Supp Table 8. List of sequences used for the ArMV population's differentiation analyses presented in supp table 7.

Supp Table 9. Measurements of genetic differentiation of grapevine fanleaf virus (GFLV) populations grouped by the geographical origin of the isolates. Sequences from the old world (Turkey, Iran, France, Hungary, Germany, Slovenia, Russia, Italy, Switzerland); new world (Canada, USA, Chile, China, South Africa); FR: France; RoTW: rest of the world other than FR; EU: Europe; Am: Americas; As: Asia; IT: Italy; SI: Slovenia; US-CA: North America; CH: Switzerland; CL: Chile and FE: far east, were analyzed. The geographic origin of isolates is specified for each sequence in Tables S2 and S3. Fixation index (F_{ST}) with associated p-value is indicated and non-significant p-values > 0.05 are underlined and in italic. Grey shading highlight F_{ST} values for far-East and Asia sequences.

Supp Table 10. List of complete nepovirus sequences submitted to GenBank after April 2020 that were not taken into account in this study.

Supp Figure 1. Unrooted maximum likelihood trees of the complete set of sequences of the genus *Nepovirus* with 110 ORF1 and 167 ORF2 nucleotide sequences. Sequences from the genera *Fabavirus* (ORF1:BBW1_NC_005289 and BBW2_KJ789136; ORF2: BBW1_NC_005290 and BBW2_KJ789137) and *Comovirus* (ORF1: CPMV_NC_003549, ORF2: CPMV_NC_003550) were used as outgroups. Colors correspond to nepovirus species affiliation with red for subgroup A sequences, blue for subgroup B sequences and green for subgroup C sequences. The scale bars show genetic distance.

Supp Figure 2. Schematic representation of intra- (triangles) and inter-species (circles) recombination events along ORF1 (left panel) and ORF2 (right panel) nucleotide sequences of representative sequences of each of the three nepovirus subgroups. Maps of the two ORFs are not drawn to scale. Numbers within triangles and circles correspond to recombination events predicted at hot spot locations. Details and precise locations of recombination events are provided in Tables S5 and S6.

Supp Figure 3. Maximum likelihood trees using all available (as of November 2019) nucleotide sequences corresponding to the coding regions 2A^{HP}, 2B^{MP} and 2C^{CP} of arabis mosaic virus (ArMV) isolates. The list of ArMV sequences is provided in Table S8. Numbers at each node indicate bootstrap values based on 100 replicates. Scale bar show genetic distance. ArMV isolates infecting monocotyledonous plants were shown in red and those infecting dicotyledonous plants are in green.

Supp Figure 4. Genetic diversity analyses of grapevine fanleaf virus (GFLV) isolates from different countries using a corpus of 40 ORF1 (left panel) and 80 ORF2 (right panel) nucleotide sequences. Colors correspond to different regions of the world with New world isolates in yellow,

Old world isolates in green, French isolates in blue and isolates from the rest of the world in red. Graphics represent π (substitution/sites) and Tajima's D (D_T) for evolution along the ORF1 and ORF2 sequences. Colored bars with # and * correspond to genetic regions being statistically valid (P-values at 0.05 and 0.001, respectively).

Supp Figure 5. Phylogenetic and diversity analyses of the 2C^{CP} coding region using a corpus of 102 nucleotide sequences of grapevine fanleaf virus (GFLV), arabis mosaic virus (ArMV) and grapevine deformation virus (GDefV) isolates with π (substitution/sites) shown along the sequence (a), phylogenetic relationships using maximum likelihood tree (b), and consensus amino acids obtained by aligning regions previously predicted and/or validate in nematode-mediated transmission (R1-R5) (c). GFLV isolates are in blue, all ArMV isolates in yellow grapevine-infecting ArMV isolates (*Vitis*-ArMV) in yellow, and non-grapevine-infecting ArMV isolates (Non-*Vitis*-ArMV) in red. For panel a, the tree CP domains (A, B, C) are delineated. For the phylogenetic tree, numbers at each node indicate bootstrap values based on 100 replicates and the scale bar shows genetic distance.

Supp Figure 6. Comparative analysis of the six possible reading frames of ORF1 from 40 grapevine fanleaf virus (GFLV) and 17 arabis mosaic virus (ArMV) sequences. Graphic representation of π (substitution/sites) along the sequence (top) and of the +1, +2, +3, -1, -2, -3 reading frames with stop codons in blue and gaps in gray (bottom). Within each reading frame that consists of 57 lines or sequences, the first 40 lines correspond to GFLV sequences, followed by nine *Vitis*-infecting ArMV sequences and finally by eight non-*Vitis*-infecting ArMV sequences. GFLV sequences are in blue, *Vitis*-infecting ArMV sequences in green and non-*Vitis*-infecting ArMV in red.

Supp Figure 7. Phylogenetic tree reconstructed from the alignment of the distal 148 and 149 amino acids of the polymerase (Pol) from grapevine fanleaf virus (GFLV) and arabis mosaic virus (ArMV) isolates. The scale bar below the tree shows genetic distance. GFLV sequences are delineated in a blue area, *Vitis*-infecting ArMV sequences in a green area and non-*Vitis*-infecting ArMV in a red area.

Supp Figure 8. Alignment of consensus amino acid sequences of the C-terminal part of the polymerase (Pol) from grapevine fanleaf virus (GFLV) isolates, *Vitis*-infected isolates of arabis mosaic virus (ArMV) and non-*Vitis*-infecting ArMV isolates. Arrows indicate identical amino acids between grapevine viruses but different from ArMV isolates infecting plants other than grapevines. The blue rectangle shows the distal 148/149 aa in the Pol domain.

ETHICS DECLARATIONS

Conflict of interest

The authors declare there are no conflicts of interest.

Ethic approval

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

LITTERATURE

1. Adams IP, Boonham N, Jones RAC (2017) First Complete Genome Sequence of Arracacha virus A Isolated from a 38-Year-Old Sample from Peru. *Genome Announcements* 5:e00141-00117
2. Bratsch S, Lockhart B, Molloy D (2017) Characterization of a New Nepovirus Causing a Leaf Mottling Disease in *Petunia hybrida*. *Plant Disease* 101:1017-1021
3. Cao M, Zhang S, Li M, Liu Y, Dong P, Li S, Kuang M, Li R, Zhou Y (2019) Discovery of Four Novel Viruses Associated with Flower Yellowing Disease of Green Sichuan Pepper (*Zanthoxylum Armatum*) by Virome Analysis. *Viruses* 11:696
4. De Souza J, Müller G, Perez W, Cuellar W, Kreuze J (2017) Complete sequence and variability of a new subgroup B nepovirus infecting potato in central Peru. *Archives of Virology* 162:885-889
5. Digiario M, Yahyaoui E, Martelli GP, Elbeaino T (2015) The sequencing of the complete genome of a Tomato black ring virus (TBRV) and of the RNA2 of three Grapevine chrome mosaic virus (GCMV) isolates from grapevine reveals the possible recombinant origin of GCMV. *Virus Genes* 50:165-171
6. Digiario M, Elbeaino T, Martelli GP (2017) Grapevine fanleaf virus and Other Old World Nepoviruses. In: Meng B, Martelli GP, Golino DA, Fuchs M (eds) *Grapevine Viruses: Molecular Biology, Diagnostics and Management*. Springer International Publishing, Cham, pp 47-82
7. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32:1792-1797
8. Elbeaino T, Digiario M, Ghebremeskel S, Martelli GP (2012) Grapevine deformation virus: Completion of the sequence and evidence on its origin from recombination events between Grapevine fanleaf virus and Arabis mosaic virus. *Virus Research* 166:136-140
9. Elbeaino T, Kiyi H, Boutarfa R, Minafra A, Martelli GP, Digiario M (2014) Phylogenetic and recombination analysis of the homing protein domain of grapevine fanleaf virus (GFLV) isolates associated with 'yellow mosaic' and 'infectious malformation' syndromes in grapevine. *Archives of Virology* 159:2757-2764
10. Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1:47-50
11. Fuchs M, Schmitt-Keichinger C, Sanfaçon H (2017) Chapter Two - A Renaissance in Nepovirus Research Provides New Insights Into Their Molecular Interface With Hosts and Vectors. In: Kielian M, Mettenleiter TC, Roossinck MJ (eds) *Advances in Virus Research*. Academic Press, pp 61-105
12. Gaafar YZA, Richert-Pöggeler KR, Sieg-Müller A, Lüddecke P, Herz K, Hartrick J, Maaß C, Ulrich R, Ziebell H (2019) Caraway yellows virus, a novel nepovirus from *Carum carvi*. *Virology Journal* 16:70

13. Gaire F, Schmitt C, Stussi-Garaud C, Pinck L, Ritzenthaler C (1999) Protein 2A of Grapevine Fanleaf Nepovirus Is Implicated in RNA2 Replication and Colocalizes to the Replication Site. *Virology* 264:25-36
14. Gao F, Lin W, Shen J, Liao F (2016) Genetic diversity and molecular evolution of arabis mosaic virus based on the CP gene sequence. *Archives of Virology* 161:1047-1051
15. Garcia S, Hily J-M, Komar V, Gertz C, Demangeat G, Lemaire O, Vigne E (2019) Detection of Multiple Variants of Grapevine Fanleaf Virus in Single Xiphinema index Nematodes. *Viruses* 11:1139
16. Ghanem-Sabanadzovic NA, Sabanadzovic S, Digiario M, Martelli GP (2005) Complete Nucleotide Sequence of the RNA-2 of Grapevine Deformation and Grapevine Anatolian Ringspot Viruses. *Virus Genes* 30:335-340
17. Gorbalenya AE, Krupovic M, Mushegian A, Kropinski AM, Siddell SG, Varsani A, Adams MJ, Davison AJ, Dutilh BE, Harrach B, Harrison RL, Junglen S, King AMQ, Knowles NJ, Lefkowitz EJ, Nibert ML, Rubino L, Sabanadzovic S, Sanfaçon H, Simmonds P, Walker PJ, Zerbini FM, Kuhn JH, International Committee on Taxonomy of Viruses Executive C (2020) The new scope of virus taxonomy: partitioning the virosphere into 15 hierarchical ranks. *Nature Microbiology* 5:668-674
18. Hily J-M, Demanèche S, Poulicard N, Tannières M, Djennane S, Beuve M, Vigne E, Demangeat G, Komar V, Gertz C, Marmonier A, Hemmer C, Vigneron S, Marais A, Candresse T, Simonet P, Lemaire O (2018) Metagenomic-based impact study of transgenic grapevine rootstock on its associated virome and soil bacteriome. *Plant Biotechnology Journal* 16:208-220
19. Hily J-M, Poulicard N, Candresse T, Vigne E, Beuve M, Renault L, Velt A, Spilmont A-S, Lemaire O (2020) Datamining, genetic diversity analyses and phylogeographic reconstructions redefine the worldwide evolutionary history of grapevine Pinot gris virus and grapevine berry inner necrosis virus. *Phytobiomes Journal* 4:165-177
20. I. Cigsar MD, K. Gokalp, N. Abou Ghanem-Sabanadzovic, A. De Stradis, D. Boscia, G.P. Martelli (2003) Grapevine deformation virus, a novel nepovirus from turkey. *Journal of Plant Pathology* 85:183-191
21. Imura Y, Oka H, Kimata K, Nasu M, Nakahama K, Maeda T (2008) Comparisons of complete RNA-2 sequences, pathological and serological features among three Japanese isolates of Arabis mosaic virus. *Virus Genes* 37:333-341
22. International Committee on Taxonomy of Viruses Executive C (2012) Family - Secoviridae. In: King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (eds) *Virus Taxonomy*. Elsevier, San Diego, pp 881-899
23. Isogai M, Tatuto N, Ujiie C, Watanabe M, Yoshikawa N (2012) Identification and characterization of blueberry latent spherical virus, a new member of subgroup C in the genus Nepovirus. *Archives of Virology* 157:297-303
24. Kis S, Salamon P, Kis V, Szittyá G (2017) Molecular characterization of a beet ringspot nepovirus isolated from *Begonia ricinifolia* in Hungary. *Archives of Virology* 162:3559-3562
25. Koloniuk I, Příbylová J, Fránová J (2018) Molecular characterization and complete genome of a novel nepovirus from red clover. *Archives of Virology* 163:1387-1389
26. Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution* 33:1870-1874
27. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular biology and evolution* 35:1547-1549
28. Le Gall O, Lanneau M, Candresse T, Dunez J (1995) The nucleotide sequence of the RNA-2 of an isolate of the English serotype of tomato black ring virus: RNA recombination in the history of nepoviruses. *Journal of General Virology* 76:1279-1283

29. Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451-1452
30. Ling R, Pate AE, Carr JP, Firth AE (2013) An essential fifth coding ORF in the sobemoviruses. *Virology* 446:397-408
31. Lu Q-Y, Wu Z-J, Xia Z-S, Xie L-H (2015) A new nepovirus identified in mulberry (*Morus alba* L.) in China. *Archives of Virology* 160:851-855
32. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B (2015) RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution* 1:1-5
33. Mekuria TA, Gutha LR, Martin RR, Naidu RA (2009) Genome Diversity and Intra- and Interspecies Recombination Events in Grapevine fanleaf virus. *Phytopathology* 99:1394-1402
34. Oliver JE, Vigne E, Fuchs M (2010) Genetic structure and molecular variability of Grapevine fanleaf virus populations. *Virus Research* 152:30-40
35. Osterbaan LJ, Choi J, Kenney J, Flasco M, Vigne E, Schmitt-Keichinger C, Rebelo AR, Cilia M, Fuchs M (2019) The Identity of a Single Residue of the RNA-dependent RNA Polymerase of Grapevine Fanleaf Virus Modulates Vein Clearing Symptoms in *Nicotiana benthamiana*. *Molecular Plant-Microbe Interactions*
36. Pagán I (2018) The diversity, evolution and epidemiology of plant viruses: A phylogenetic view. *Infection, Genetics and Evolution* 65:187-199
37. Pompe-Novak M, Gutiérrez-Aguirre I, Vojvoda J, Blas M, Tomažič I, Vigne E, Fuchs M, Ravnikar M, Petrovič N (2007) Genetic variability within RNA2 of Grapevine fanleaf virus. *European Journal of Plant Pathology* 117:307-312
38. Rebenstorf K, Candresse T, Dulucq MJ, Büttner C, Obermeier C (2006) Host species-dependent population structure of a pollen-borne plant virus, Cherry leaf roll virus. *Journal of virology* 80:2453-2462
39. Rivera L, Zamorano A, Fiore N (2016) Genetic divergence of tomato ringspot virus. *Archives of Virology* 161:1395-1399
40. Roberts JMK, Anderson DL, Durr PA (2018) Metagenomic analysis of Varroa-free Australian honey bees (*Apis mellifera*) shows a diverse Picornavirales virome. *Journal of General Virology* 99:818-826
41. Rowhani A, Daubert SD, Uyemoto JK, Al Rwahnih M, Fuchs M (2017) American Nepoviruses. In: Meng B, Martelli GP, Golino DA, Fuchs M (eds) *Grapevine Viruses: Molecular Biology, Diagnostics and Management*. Springer International Publishing, Cham, pp 109-126
42. Sanfaçon H (2008) Nepovirus. In: Mahy BWJ, Van Regenmortel MHV (eds) *Encyclopedia of Virology (Third Edition)*. Academic Press, Oxford, pp 405-413
43. Sanfaçon H (2015) Secoviridae: A Family of Plant Picorna-Like Viruses with Monopartite or Bipartite Genomes. eLS. John Wiley & Sons, Ltd
44. Sanfaçon H, Dasgupta I, Fuchs M, Karasev AV, Petrzik K, Thompson JR, Tzanetakis I, van der Vlugt R, Wetzel T, Yoshikawa N (2020) Proposed revision of the family Secoviridae taxonomy to create three subgenera, “Satsumavirus”, “Stramovirus” and “Cholivirus”, in the genus Sadwavirus. *Archives of Virology* 165:527-533
45. Schellenberger P, Andret-Link P, Schmitt-Keichinger C, Bergdoll M, Marmonier A, Vigne E, Lemaire O, Fuchs M, Demangeat G, Ritzenthaler C (2010) A stretch of 11 amino acids in the betaB-betaC loop of the coat protein of grapevine fanleaf virus is essential for transmission by the nematode *Xiphinema index*. *Journal of virology* 84:7924-7933
46. Simmonds P, Aiewsakun P (2018) Virus classification – where do you draw the line? *Archives of Virology* 163:2037-2046
47. Sokhandan-Bashir N, Melcher U (2012) Population genetic analysis of grapevine fanleaf virus. *Archives of Virology* 157:1919-1929

48. Sorrentino R, De Stradis A, Russo M, Alioto D, Rubino L (2013) Characterization of a putative novel nepovirus from *Aeonium* sp. *Virus Research* 177:217-221
49. Thompson JR, Kamath N, Perry KL (2014) An Evolutionary Analysis of the Secoviridae Family of Viruses. *PLOS ONE* 9:e106305
50. Vigne E, Bergdoll M, Guyader S, Fuchs M (2004) Population structure and genetic variability within isolates of Grapevine fanleaf virus from a naturally infected vineyard in France: evidence for mixed infection and recombination. *Journal of General Virology* 85:2435-2445
51. Vigne E, Demangeat G, Komar V, Fuchs M (2005) Characterization of a naturally occurring recombinant isolate of Grapevine fanleaf virus. *Archives of Virology* 150:2241-2255
52. Vigne E, Marmonier A, Fuchs M (2008) Multiple interspecies recombination events within RNA2 of Grapevine fanleaf virus and *Arabis* mosaic virus. *Archives of Virology* 153:1771-1776
53. Vigne E, Garcia S, Komar V, Lemaire O, Hily J-M (2018) Comparison of Serological and Molecular Methods With High-Throughput Sequencing for the Detection and Quantification of Grapevine Fanleaf Virus in Vineyard Samples. *Frontiers in Microbiology* 22:2726
54. Walker M, Chisholm J, Wei T, Ghoshal B, Saeed H, Rott M, Sanfaçon H (2015) Complete genome sequence of three tomato ringspot virus isolates: evidence for reassortment and recombination. *Archives of Virology* 160:543-547
55. Wetzel T, Fuchs M, Bobko M, Krczal G (2002) Size and sequence variability of the *Arabis* mosaic virus protein 2A. *Archives of Virology* 147:1643-1653
56. Yasmin T, Nelson BD, Hobbs HA, McCoppin NK, Lambert KN, Domier LL (2017) Molecular characterization of a new soybean-infecting member of the genus *Nepovirus* identified by high-throughput sequencing. *Archives of Virology* 162:1089-1092

Figures

Figure 1

SubGroup A
SubGroup B
SubGroup C

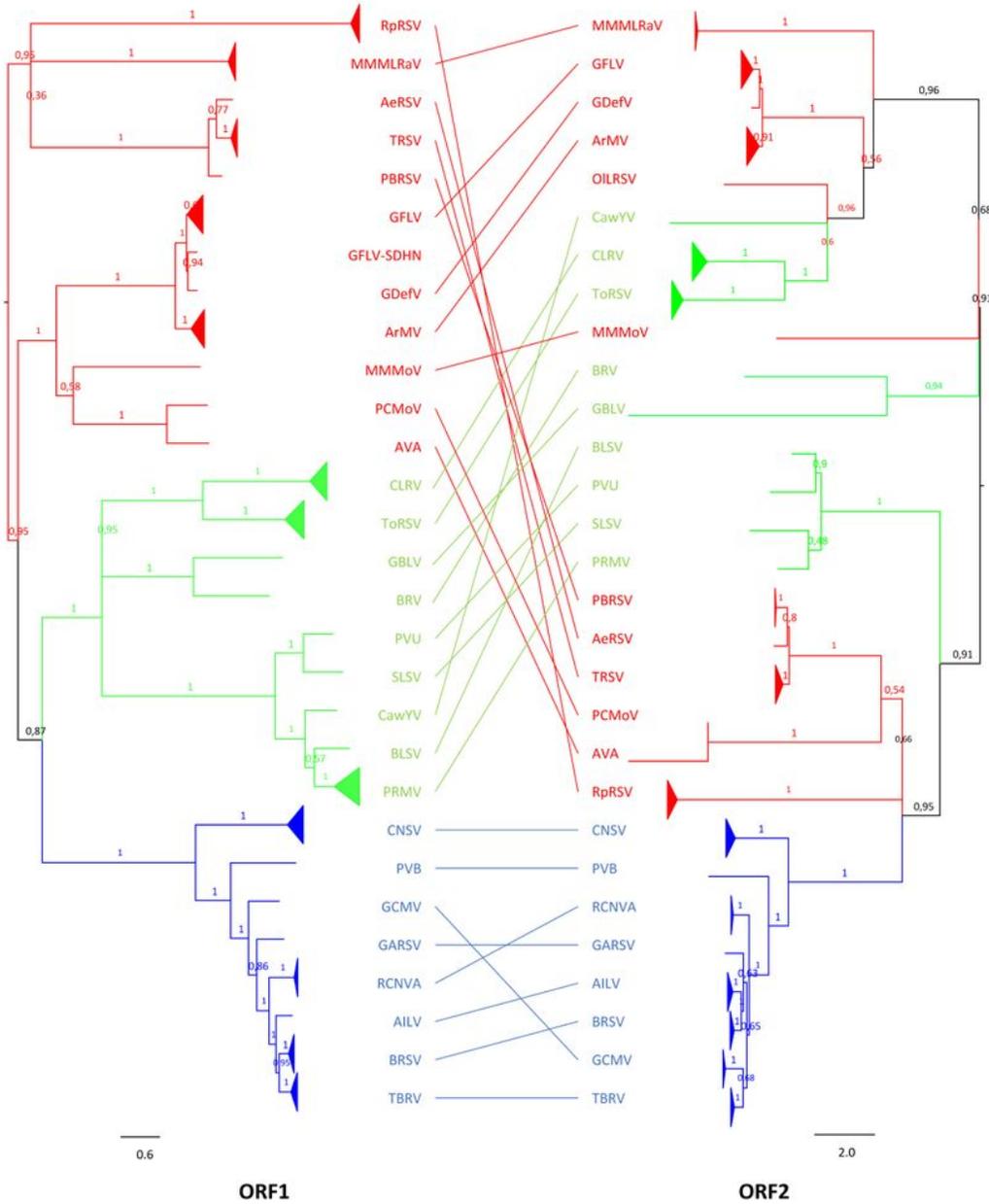


Figure 1

Tanglegram of maximum likelihood phylogenetic trees inferred from 110 ORF1 and 167 ORF2 nucleotide sequences of nepoviruses. Colors correspond to species affiliation to one of the three nepovirus subgroups: red for subgroup A sequences; blue for subgroup B sequences; and green for subgroup C

sequences. Clades with several sequences from the same species are collapsed. Numbers at nodes indicate bootstrap values based on 100 replicates. The scale bar corresponds to the number of substitutions per site.

Figure 2

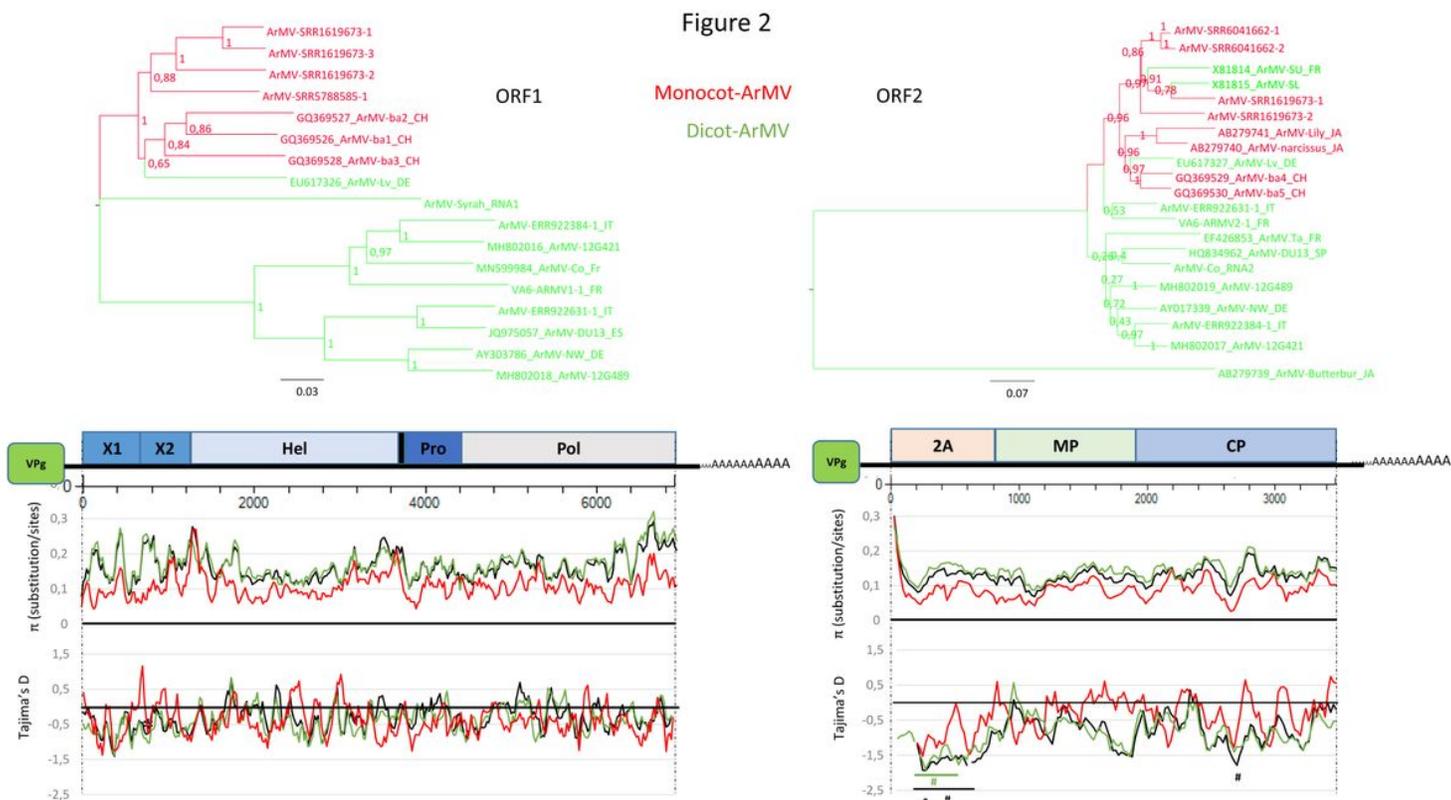


Figure 2

Phylogenetic and diversity analyses of arabis mosaic virus (ArMV) isolates from different plants using a corpus of 17 ORF1 (left panel) and 21 ORF2 (right panel) nucleotide sequences. Colors correspond to hosts with red for monocotyledonous plants, green for dicotyledonous plants and black for all the plants. Maximum likelihood phylogenetic trees are shown. Numbers at each node indicate bootstrap values based on 100 replicates and scale bars show genetic distance. Graphics represent π (substitution/sites) and Tajima's D (DT) for evolution along the ORF1 and ORF2 sequences. Colored bars with # and * correspond to statistically validated regions (P-values at 0.05 and 0.001, respectively).

Figure 3

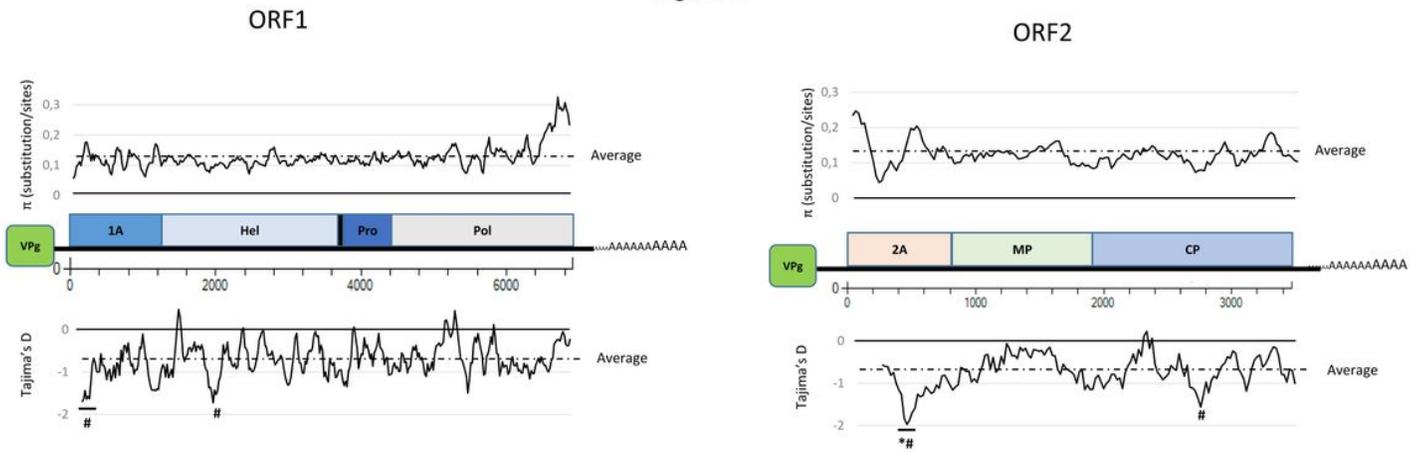


Figure 3

Genetic diversity analyses of grapevine fanleaf virus (GFLV) isolates using a corpus of 40 ORF1 (left panel) and 80 ORF2 (right panel) nucleotide sequences. Graphics represent π (substitution/sites) and Tajima's D (DT) for evolution along the ORF1 and ORF2 sequences. Bars and # and * correspond to statistically validated regions (P-values at 0.05 and 0.001, respectively).

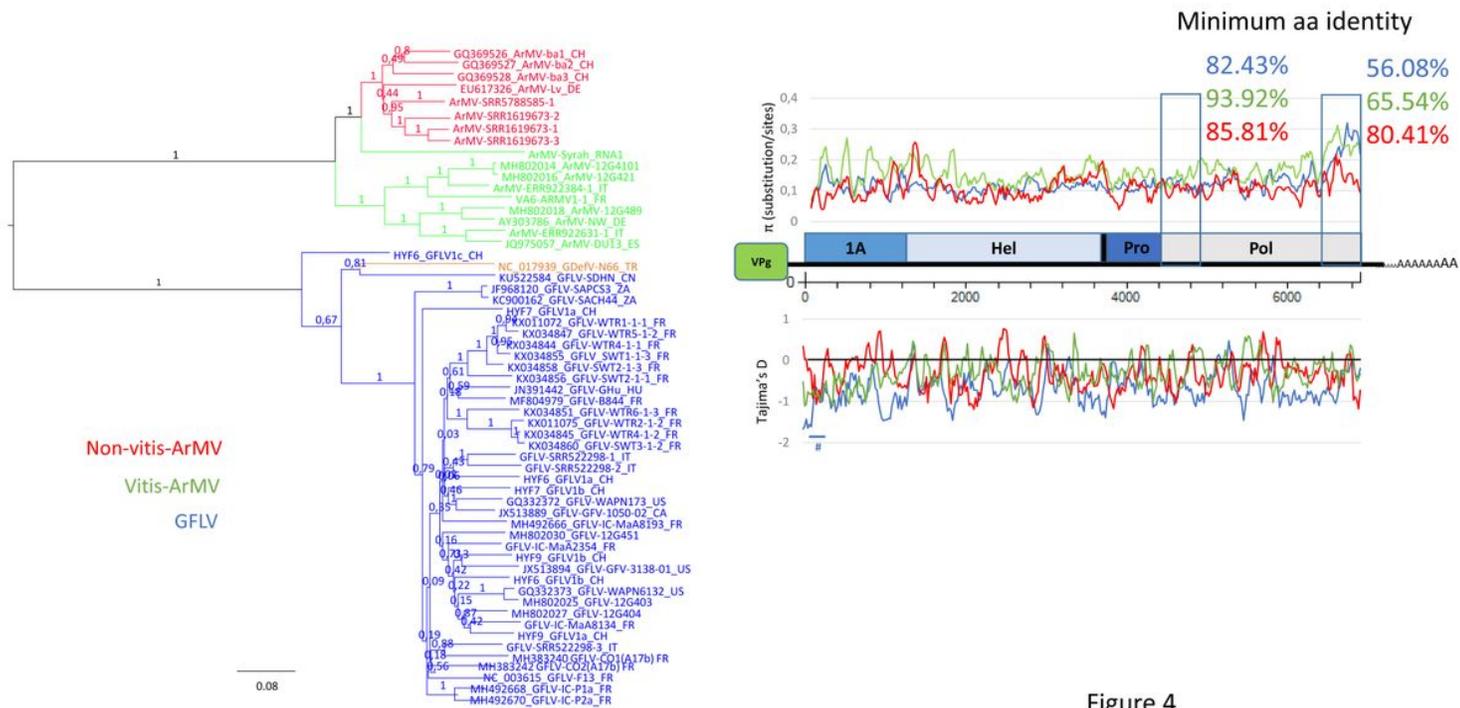


Figure 4

Figure 4

Phylogenetic and diversity analyses of grapevine fanleaf virus (GFLV), arabis mosaic virus (ArMV) and grapevine deformation virus (GDefV) using a corpus of 58 ORF1 nucleotide sequences. GFLV isolates are in blue, GDefV isolates in peach, ArMV isolates from grapevines (Vitis-ArMV) in green and ArMV isolates from other plants (non-vitis ArMV) in red. Maximum likelihood phylogenetic trees are shown. Numbers at each node indicate bootstrap values based on 100 replicates and the scale bar shows genetic distance. Graphics represent π (substitution/sites) and Tajima's D (DT) for evolution along the ORF1 sequence. Percentages correspond to minimum aa identity for both framed regions within the polymerase domain. The colored bar with # corresponds to a statistically validated region at 0.05.

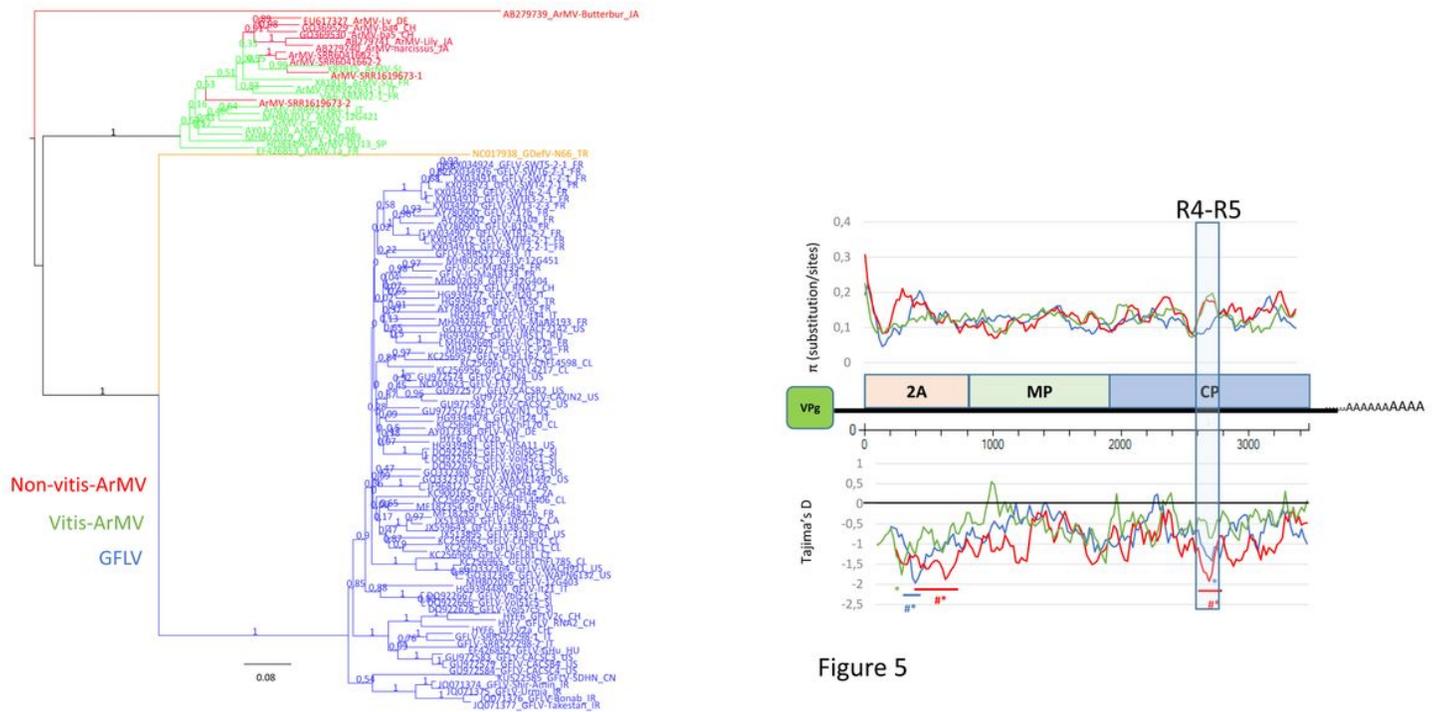


Figure 5

Phylogenetic and genetic diversity analyses of grapevine fanleaf virus (GFLV), arabis mosaic virus (ArMV) and grapevine deformation virus (GDefV) using a corpus of 102 ORF2 nucleotide sequences. GFLV isolates are in blue, GDefV in peach, ArMV isolates from grapevines (Vitis-ArMV) in green and ArMV isolates from other plants (non-vitis ArMV) in red. Maximum likelihood phylogenetic trees are shown. Numbers at each node indicate bootstrap values based on 100 replicates and the scale bar shows genetic distance. Graphics represent π (substitution/sites) and Tajima's D (DT) for evolution along the ORF2 sequence. The boxed area corresponds to the R4-R5 region of the 2CCP domain. Colored bars with # and * correspond to statistically validated regions (P-values at 0.05 and 0.001, respectively).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Tables15S1S4S9S10Nepoviruses.pdf](#)
- [ArchVirolNepovirusesNCBIsubmitted20201217.fa](#)
- [TableS5S6.pdf](#)
- [TableS3S8.pdf](#)
- [TableS2S7.pdf](#)
- [s1.jpg](#)

- [s2.jpg](#)
- [s3.jpg](#)
- [s4.jpg](#)
- [s5.jpg](#)
- [s6.jpg](#)
- [s7.jpg](#)
- [s8.jpg](#)