bioRxiv preprint doi: https://doi.org/10.1101/314674; this version posted August 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Grapevine virus L: a Novel Vitivirus in Grapevine

Humberto Debat[1Ω*], Diego Zavallo[2Ω], Reid Soltero Brisbane[3], Darko Vončina[4], Rodrigo P.P. Almeida[5], Arnaud G. Blouin [6], Maher Al Rwahnih[7], Sebastian Gomez-Talquenca[8*] and Sebastian Asurmendi[2,9]

[1]Instituto de Patología Vegetal, Centro de Investigaciones Agropecuarias, Instituto Nacional de Tecnología Agropecuaria (IPAVE-CIAP-INTA), Córdoba, Argentina, X5020ICA.

[2]Instituto de Biotecnología, Centro de Investigación en Ciencias Veterinarias y Agronómicas (IB-CICVyA-INTA), Buenos Aires, Argentina, 1686.

[3]Foundation Plant Services, Davis, CA 95616, USA

[4]Department of Plant Pathology, Faculty of Agriculture, University of Zagreb, Zagreb, Croatia

[5]Department of Environmental Science, Policy and Management, University of California, Berkeley, CA, USA.

[6]The New Zealand Institute for Plant & Food Research Limited, Private Bag 92169, Auckland, 1142, New Zealand

[7]Department of Plant Pathology, University of California, Davis, CA, USA

[8]Estación Experimental Agropecuaria Mendoza, Instituto Nacional de Tecnología Agropecuaria (EEA-Mendoza-INTA), Luján de Cuyo, Mendoza, Argentina, 5534.

[9]CONICET, Argentina.

[Ω] These authors contributed equally to this work

ORCID ID:

HD, 0000-0003-3056-3739; DZ, 0000-0002-9021-2175; DV, 0000-0001-6093-5471; RPPA, 0000-0003-2888-9617; AGB, 0000-0003-1360-1529; SGT, 0000-0001-9265-8333; SA, 0000-0001-9516-5948

* Correspondence:

Humberto Debat, debat.humberto@inta.gob.ar

Sebastian Gomez-Talquenca, gomez.talquenca@inta.gob.ar

bioRxiv preprint doi: https://doi.org/10.1101/314674; this version posted August 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Vitivirus are ssRNA(+) viruses in the family *Betaflexiviridae* (subfamily *Trivirinae*). There are currently ten ICTV recognized virus species in the genus; nevertheless, the extended use of NGS technologies is rapidly expanding their diversity and six more have been proposed recently. Here, we present the characterization of a novel virus from grapevines, which fits the genomic architecture and evolutionary constraints to be classifiable within the *Vitivirus* genus. The detected virus sequence is 7,607 nt long, including a typical genome organization of ORFs encoding a replicase (RP), a 22 kDa protein, a movement protein, a coat protein (CP) and a nucleic acid binding protein. Here, we present the characterization of a novel virus from grapevines. Phylogenetic analyses based on the predicted RP and CP protein unequivocally places the new virus within the *Vitivirus* genus. Multiple independent RNAseq data confirmed the presence of the detected virus in berries at diverse developmental stages. Additionally, we detected, confirmed, and assembled virus sequences from grapevine samples of distinct cultivars from America, Europe, Asia and Oceania, sharing 74.9%-97.9% nt identity, suggesting that the identified virus is widely distributed and diverse. We propose the name grapevine virus L (GVL) to the detected *Vitivirus*.

Keywords

Vitivirus, Grapevine, virus discovery, *Betaflexiviridae*

Accession numbers

Grapevine virus L sequences have been deposited in GenBank under accession numbers: MH248020 (GVL-RI), MH643739 (GVL-KA), MH681991 (GVL-VL), MH686191 (GVL-SB)

bioRxiv preprint doi: https://doi.org/10.1101/314674; this version posted August 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Vitiviruses have flexuous, non-enveloped, filamentous virus particles of 725-785 nm and 12 nm in length and diameter, respectively, with a nucleocapsid that is cross-banded and diagonally striated. Vitiviruses have a linear ssRNA(+) genome (~7.3-7.6 kb), with a methylated nucleotide cap at the 5' end and a 3' poly (A) tail [1-2]. There are ten species of vitiviruses recognized by the ICTV, nevertheless, six new species have been proposed recently, five of those infecting grapevine (*Vitis vinifera*) [3-7]. Grapevine is the most prevalent natural host of vitiviruses, but they have also been found to infect several crops such as mint (*Mentha x glaciaris*), arracacha (*Arracacia xanthorrhiza*) blue agave (*Agave tequilana*), kiwi (*Actinidia chinense*) and blackberry (Rubus spp) [8-11]. Vitiviruses appear to be latent in *V. vinífera* cultivars, and so far, only *Grapevine virus A* and *Grapevine virus B* have been consistently associated to grapevine diseases of the rugose wood complex (Grapevine vitiviruses) or Shiraz disease (reviewed by Minafra [12]). The etiological role of the recently described vitiviruses should be assessed in order to establish its relationships with known or unknown viral diseases. In addition, the synergistic effects between vitiviruses and other grapevine viruses appears to be significant [13]. The availability of complete sequences of these viruses could allow the development of full-length infectious clones, fulfillment of the Koch's postulates, and improve our understanding of the biological role of these viruses [14]. Here, we present the characterization of a novel vitivirus for which we tentatively propose as grapevine virus L (GVL). We were able to detect and characterize GVL in several distinctive *Vitis vinifera* cultivars from four continents.

Material and Methods

Virus discovery, confirmation, and annotation were implemented as described in [4-5, 15-16]. RT-PCR assays were deployed as reported in Al Rwahnih et al. [17]. Raw RNA data from several Sequence Read Archive (SRA) accessions was downloaded from the National Center for Biotechnology Information database (NCBI). These files was processed in the following form: trimmed and filtered with the Trimmomatic tool as implemented in http://www.usadellab.org/cms/?page=trimmomatic, and the resulting reads of each library were assembled *de novo* with Trinity v2.6.6 release with standard parameters [18]. The obtained transcripts were subjected to bulk local blastx searches (E-value < 1e-5) against a refseq virus protein database available at ftp://ftp.ncbi.nlm.nih.gov/refseq/release/viral/viral.1.protein.faa.gz. The resulting hits were explored manually and curated by iterative mapping using Bowtie2 available at http://bowtie-bio.sourceforge.net/bowtie2/. ORFs were predicted by ORFfinder (https://www.ncbi.nlm.nih.gov/orffinder/) and annotated with the NCBI conserved domain search tool as implemented in https://www.ncbi.nlm.nih.gov/Structure/cdd. Phylogenetic insights were generated by multiple amino acid alignments of replicase proteins (BLOSUM62 scoring matrix) using as best-fit algorithm E-INS-i, which drives local alignment with generalized affine gap costs, and thus is applicable for RNA polymerases where diverse domains are dispersed among several highly divergent regions, by MAFTT v7.394 as implemented in https://mafft.cbrc.jp/alignment/software/ . The aligned proteins were subsequently used as input for FastTree 2.1.5 http://www.microbesonline.org/fasttree/ maximum likelihood phylogenetic trees (best-fit model = JTT-Jones-Taylor-Thorton with single rate of evolution for each site = CAT) computing local support values with the Shimodaira-Hasegawa test (SH) and 1,000 resamples. SNPs were detected by the FreeBayes v0.9.18S tool with standard parameters implemented in https://github.com/ekg/freebayes.

bioRxiv preprint doi: https://doi.org/10.1101/314674; this version posted August 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Results and Discussion

In order to explore the potential grapevine viral landscape on High Throughput Sequencing (HTS) publicly available libraries, we downloaded raw RNA data from several SRA accessions on the NCBI database, finding known virus in the vast majority of tested samples. The raw data were assembled *de novo* and the obtained transcripts were subjected to bulk local blastx searches (E-value < 1e-5) against a refseq virus protein database. The resulting hits were explored manually. One dataset, specifically NCBI Bioproject PRJNA400621, SRX3144921-SRX3144956, composed of RNA libraries from grape berry samples of *V. vinifera* cv. Riesling and cv. Cabernet sauvignon from Beijing, China [19], presented a 6,155 nt long transcript from the assembled transcriptome of the SRX3144956 library. This sequence obtained a highly significant hit (E-value = 0; sequence identity 69%) with the replicase protein of *Grapevine virus E* (GVE, YP_002117775.1, [20]). Iterative read mapping allowed to extend and polish the assembled contig into a 7,607 nt long sequence, supported by 25,112 reads and a mean coverage of 161.7X, sharing the genome organization of vitiviruses (**Figure 1.A**; **Table 1**). The assembled virus sequence has a pairwise identity with GVE of 68.2% at the genomic RNA level. The predicted structure of the detected virus is conformed by a 67 nt 5´UTR, followed by five ORFs and a 66 nt 3´UTR, excluding the poly (A) tail of at least three adenosines, and unknown length. Both predicted UTRs are highly similar in length and size in relation to other *Vitivirus*, more precisely GVE and the proposed GVI and GVG (**Supp. Figure 1**). ORF1 (67-5,158 nt coordinates) encodes a replicase protein (RP) of 1,696 aa (192.1 kDa), sharing a 70.2% with GVE replicase (NC_011106). We employed the NCBI conserved domain search tool to annotate the RP, which presents a characteristic domain architecture of *Vitivirus*. The 5' region of RP presents a viral methyltransferase domain (E-value = 1.76e-56; Pfam = pfam01660; 46-337 aa coordinates), which has been involved in mRNA capping, followed by a DEXDc type viral RNA helicase (E-value = 1.47e-12; Pfam = pfam01443; 750-1047 aa coordinates). Interestingly, an Alkylated DNA repair dioxygenase domain (AlkB), of the2OG-Fe(II) oxygenase superfamily, was identified within the helicase domain (E-value = 2.34e-06; COG = COG3145; 917-1032 aa coordinates) as previously described for an American isolate of GVE [21]. Finally, as expected, an RNA dependent RNA polymerase domain was found at the 3' region of the RP (RdRP_2; E-value = 3.47e-27; Pfam = pfam00978; 1313-1637 aa coordinates).

Although vitiviruses utilize a series of sub-genomic RNAs (sgRNA) as general strategy for expression of ORF2-5, it is worth mentioning that ORF2 (5,155-5,766 nt) of the detected virus overlaps in the tetranucleotide AUGA with ORF1. This overlapping cue might probably function as a termination–reinitiation strategy signal, which was determined for caliciviruses and victoriviruses for coupled start-stop translation [22]. This putative AUGA signal between ORF1-ORF2, although not mentioned, is present in the genome sequence of the recently proposed grapevine virus J [5]. The predicted ORF2 encodes a leucine rich (15.3% L) 22.8 kDa protein, of unknown function, which is the most highly divergent protein of the identified virus (44% similarity and 98% coverage to the ORF2 encoded hypothetical protein of GVE, AHB08905). ORF3 (5,795-6,607 nt) encodes a putative movement protein (MP) of 270 aa (29.7 kDa) with a viral movement protein domain (E-value = 6.61e-07; Pfam = pfam01107) between the 26-195 aa coordinates. ORF4 (6,522-7,121 nt), partially overlapping with ORF3, encodes a coat protein (CP) 199 aa long (21.4 kDa) presenting a *Trichovirus* coat protein-like domain (Tricho_coat; E-value = 4.26e-45; Pfam = pfam05892; 37-199 aa coordinates). The predicted CP of the identified virus shares 78.4% similarity with GVE-CP. Lastly, ORF5 (7,139-7,540 nt)

bioRxiv preprint doi: https://doi.org/10.1101/314674; this version posted August 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

encodes a 133 aa putative nucleic acid binding protein (15 kDa) with a nucleic acid binding domain (Viral_NABP; E-value = 1.40e-06; Pfam = pfam05515) between the 5-76 aa coordinates. This region shares significant similarity with the CTV_P23 domain (E-value =3.79e-06) of *Citrus tristeza virus* (*Closteroviridae*) P23 protein which is involved in asymmetrical RNA Accumulation [23].

The criteria for species demarcation in the family *Betaflexiviridae*, proposed by ICTV, is less than ca. 72% overall nucleotide identity and below 80% amino acid identity of CP or replicase predicted proteins [1]. Given that the detected virus shares 68.2% similarity with GVE at the genomic RNA level, and a 70.2% and 78.2% identity with the respective RP and CP proteins (**Supp. Figure 2**), is tentatively suggested that the detected virus sequence corresponds to a new species, for which we propose the name grapevine virus L (GVL). In order to entertain this hypothesis, based on genetic criteria, we gathered evolutionary insights of the putative GVL sequence identified in cv. Riesling grapevines from China (GVL-RI) in the context of the *Betaflexiviridae* family. We generated a phylogenetic tree based on RP proteins corresponding to the 93 refseq *Betaflexiviridae* RP sequences available at NCBI. Unequivocally, the putative GVL RP clustered within the *Vitivirus* genus (**Supp. Figure 3**). To provide local tree topology and additional phylogenetic insights we generated maximum likelihood phylogenetic trees of both the RP and CP protein of accepted and proposed members of the *Vitivirus* genus. In both cases, putative GVL clustered within one of the two major sub-clades, conformed by GVE and the proposed species GVG, GVI, and branching more distantly, the agave tequilana leaf virus (**Figure 1.B-C**). These results support the tentative assignment of the detected virus as a probable member of the *Vitivirus* genus.

To validate the putative GVL detection in SRX3144956 library, we took advantage of the data generated by Chen et al [19]. We mapped the 1,014,833,524 raw reads available at the PRJNA400621 project to the assembled GVL sequence, using as threshold two nt mismatches and a 22 nt seed and a cut-off value >10 reads hits per library with the Bowtie2 tool. None of the 18 RNA libraries of cv. Cabernet Sauvignon had the putative GVL. Interestingly, we were able to detect the virus in six additional cv Riesling libraries of the same study, obtained from independent biological samples conformed by grape berries sampled in three distinct developmental stages (E-L 35, E-L 36 and E-L 38) (**Supp. Table 1**). Moreover, while analyzing the intrinsic variability of the detected GVL sequences, we noticed that two of the samples presented 24 highly supported SNPs at frequencies >0.5 as detected by the FreeBayes v0.9.18S tool (**Supp. Table 2**). Given the low p-values of the variants and that 70% of the detected SNPs were silent, we speculate that these SNPs are not artifactual.

In order to explore the distribution of GVL, we gathered data from grapevine samples from Croatia [24], USA, and New Zealand. A collection of HTS libraries obtained from petioles RNA samples of autochthonous Croatian grapevines (cv Babica, cv Vlaska, cv Dobricic and cv Ljutun; Bioproject PRJNA415169) had virus RNA reads similar to GVL. In turn, *de novo* assembly and curation of the putative GVL transcript by iterative mapping resulted in a 7,584 nt virus genome (GVL-VL, mean coverage of 675X), sharing the genomic organization and a 75.1 % nt identity with GVL-RI (**Supp. Table 3**). The presence of this virus was subsequently confirmed by targeted RT-PCR (see below) of the HTS sequenced samples. Two PCR products were bi-directionally Sanger-sequenced and the obtained sequences shared a 100 % identity with GVL-VL.

The sample from USA corresponded to a quarantine selection Katelin (KA), which was received in 2012 for inclusion in the Foundation Plant Services (FPS, UC-Davis, CA) collection. The vine was grown in a screenhouse and assayed for known grapevine viruses as described previously [16]. In addition, total RNA from

bioRxiv preprint doi: https://doi.org/10.1101/314674; this version posted August 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the positive sample was high-throughput sequenced and the obtained ca. 11.5 million raw reads were filtered and *de novo* assembled with parameters described in Al Rwahnih et al. 2018 [16]. A contig with high similarity with GVL was identified by BLAST and subsequently refined, resulting in a 7,591 nt virus draft genome (GVL-KA, MH643739), sharing the genomic architecture and a 97.9% nt identity with GVL-RI (**Supp. Table 3**).

Lastly, the presence of GVL was confirmed in an HTS library derived from total RNA isolated from cv. Sauvignon Blanc, which was originally collected in April 2016 in the region of Marlborough, South Island, New Zealand. The virus was detected in only one of the 225 Vitis sp. surveyed by HTS. The presence of the virus was confirmed by RT-PCR and Sanger sequencing. A draft genome with a partial truncation at 5' was assembled, encompassing 7,365 nt, lacking both the 5'UTR and a ca. 184 nt coding region of the replicase (GVL-SB, mean coverage of 166X). The obtained sequence shared a 75.2% nt identity with GVL-RI. All draft genome sequences were annotated both structurally and functionally, showing consistent signatures of vitiviruses and a shared genome organization (**Supp. Figure 4**). In turn, sequence identity values were calculated indicating high sequence similarity in the predicted coat protein (92.5% to 99.5% aa identity) and significant divergence in the 22 kDa predicted protein (53.2% to 96.6% aa identity) (**Supp. Table 3**).

Further, an RT-PCR assay was deployed as reported in Al Rwahnih et al. [17] to investigate the prevalence of GVL in grapevine samples from USA, using forward [AGTTGAAGTCTAGGTGCACAC] and reverse [GTACTCAGACTTCCCCGATCTA] primers designed to target the MH248020 sequence of GVL-RI. The specificity of this primer pair in the detection of GVL was confirmed by RT-PCR using Total nucleic acid (TNA) from vines infected with GVA, GVB, GVD, GVE, GVH, GVG, GVJ. TNA was freshly extracted from leaf petiole tissue using the MagMax 96 Viral RNA isolation kit with the MagMax Express-96 magnetic particle processor (Thermo Fisher Scientific, USA). Using the above primers in those assays, none of the grapevines that were infected with the non-target vitiviruses were found to be positive, whereas a sample of grapevine cv. Katelin tested positive. The development of this specific RT-PCR assay will allow for the detection of GVL in field tests, and in clean-stock programs, facilitating a more effective control of this virus.

Given the intrinsic significant variability observed among the diverse GVL identified isolates, in the context of an overall interspecies affinity of GVL and GVE, we proceeded with additional phylogenetic analyses. We integrated the predicted products of four assembled GVL sequences and the five available, tentatively complete, genome sequences of GVE in multiple amino acid alignments (isolates TvAQ7, of *V. labruscana* from Japan; WAHH2, of *V. vinifera* cv Cabernet sauvignon from USA; SA94, of *V. vinifera* cv. Shiraz from South Africa; VVL-101, *V. vinifera* cv. Vlaska from Croatia and GFMG-1 from China). The maximum likelihood phylogenetic trees based both on RP and CP alignments mirrored the preceding tree topology presented in **Figure 1.B** and supported clade branching between GVL and GVE in the context of clustering within *Vitiviruses* (**Figure 2.A-B**). The availability of more RNA sequences corresponding to GVL would provide insights into the evolutionary trajectory and phylodynamics of GVL. The simultaneous detection of GVL in the Americas, Europe, Asia and Oceania in diverse *V. vinifera* cultivars suggest that this novel virus is widely distributed but not widely spread in any of the environment surveyed. The emerging biological properties of GVL remain elusive. Given that the samples harboring GVL also presented several grapevine viruses and viroids, including *Grapevine rupestris stem pitting-associated virus* (*Foveavirus*), *Grapevine Syrah virus-1* (*Marafivirus*), *Grapevine leafroll-associated virus 3* (*Ampelovirus*), *Grapevine yellow speckle viroid* 1

bioRxiv preprint doi: https://doi.org/10.1101/314674; this version posted August 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

& 2 (*Apscaviroid*), and *Hop stunt viroid* (*Hostuviroid*), future studies should unravel whether GVL presence is associated to any specific symptoms in grapevine.

bioRxiv preprint doi: https://doi.org/10.1101/314674; this version posted August 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Acknowledgements

This work was supported by project PE1131022 of the Instituto Nacional de Tecnología Agropecuaria (INTA) and by ANPCyT PICT 2015-1532 and PICT 2016-0429. The funders had no role in study design and analysis, decision to publish, or preparation of the manuscript. The research from New Zealand is funded by New Zealand Winegrowers Inc. and The Ministry of Business, Innovation & Employment. Many thanks to Alfredo Diaz Lara, Department of Plant Pathology, University of California, Davis, CA for his technical assistance.

Author contributions

HD, SGT and SA designed the study. RSB, DV, RPPA, AGB and MAR, performed experiments, generated and analyzed data. HD integrated the data together with SA. HD wrote the initial draft of the manuscript and DZ, SGT, AGB, RSB, DV, RPPA, MAR and SA revised the manuscript. All authors approved the final version.

Conflict of interest

The authors declare that they have no conflict of interest.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors. Therefore, informed consent was not required for this work.

bioRxiv preprint doi: https://doi.org/10.1101/314674; this version posted August 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. Adams MJ, Candresse T, Hammond J, Kreuze JF, Martelli GP, Namba S, Pearson MN, Ryu KH, Saldarelli P, Yoshikawa N (2012) Family Betaflexiviridae. In: Virus Taxonomy – Ninth Report on the International Committee on Taxonomy of Viruses. AMQ King, MJ Adams, EB Carstens, EJ Lefkowitz, eds. Elsevier Academic Press, London, pp 920-941

2. Du Preez J, Stephan D, Mawassi M, Burger JT (2011) The grapevine-infecting vitiviruses, with particular reference to grapevine virus A. Arch Virol. 156:1495.

3. Blouin AG, Chooi KM, Warren B, Napier KR, Barrero RA, MacDiarmid RM (2018) Grapevine virus I, a putative new vitivirus detected in co-infection with grapevine virus G in New Zealand. Arch Virol. 163:1371-1374.

4. Blouin AG, Keenan S, Napier KR, Barrero RA, MacDiarmid RM (2018) Identification of a novel vitivirus from grapevines in New Zealand. Arch Virol.163:281–284.

5. Diaz-Lara A, Golino D, Al Rwahnih M (2018) Genomic characterization of grapevine virus J, a novel virus identified in grapevine. Arch Virol. https://doi.org/10.1007/s00705-018-3793-y

6. Candresse T, Theil S, Faure C, Marais A (2018) Determination of the complete genomic sequence of grapevine virus H, a novel vitivirus infecting grapevine. Arch Virol. 163:277–280.

7. Jo Y, Song MK, Choi H, Park JS, Lee JW, Lian S, Cho WK (2017) Genome Sequence of Grapevine Virus K, a Novel Vitivirus Infecting Grapevine. Genome Announc. 5:e00994-17.

8. Tzanetakis IE, Postman JD, Martin RR (2007) Identification, detection and transmission of a new vitivirus from Mentha. Arch Virol.152:2027–2033.

9. Oliveira LM, Orílio AF, Inoue-Nagata AK, Nagata T, Blawid R (2017) A novel vitivirus-like sequence found in Arracacia xanthorrhiza plants by high throughput sequencing. Arch Virol. 162: 2141-2144.

10. Blouin AG, Chavan RR, Pearson MN, Macdiarmid RM, Cohen D (2012) Detection and characterisation of two novel vitiviruses infecting Actinidia. Arch Virol. 157:713–722.

11. Hassan M, Shahid MS, Tzanetakis IE. (2018). Molecular characterization and detection of a novel vitivirus infecting blackberry. Arch Virol. https://doi.org/10.1007/s00705-018-3931-6

12. Minafra A, Mawassi M, Goszczynski D, Saldarelli P (2017) Grapevine Vitiviruses. In: Grapevine Viruses. Meng B, Martelli GP, Golino DA, Fuchs M. eds. Springer International Publishing AG.

13. Rowhani A, Daubert S, Arnold K, Al Rwahnih M, Klaassen V, Golino D, Uyemoto JK (2018) Synergy between grapevine vitiviruses and grapevine leafroll viruses. Eur J Plant Pathol. https://doi.org/10.1007/s10658-018-1426-7

14. Goszczynski DE (2015) Brief report of the construction of infectious DNA clones of South African genetic variants of Grapevine virus A and Grapevine virus B. Springer Plus. 4:739.

15. Al Rwahnih M, Daubert S, Golino D, Islas C, Rowhani A (2015). Comparison of next-generation sequencing versus biological indexing for the optimal detection of viral pathogens in grapevine. Phytopathology. 105:758-763.

16. Al Rwahnih M, Rowhani A, Westrick N, Stevens K, Diaz-Lara A, Trouillas FP, Preece J, Kallsen C, Farrar K, Golino D (2018). Discovery of viruses and virus-like pathogens in pistachio using high-throughput sequencing. Plant Dis. 102:1419-1425.

17. Al Rwahnih M, Daubert S, Islas C, Golino DA, Rowhani A (2014). Characterization of a fifth Vitivirus in grapevine. J Plant Pathol. 96:2019-222.

bioRxiv preprint doi: https://doi.org/10.1101/314674; this version posted August 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

18. Haas B J, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, MacManes MD (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 8:1494.

19. Chen WK, Bai XJ, Cao MM, Cheng G, Cao XJ, Guo RR, Duan CQ (2017) Dissecting the Variations of Ripening Progression and Flavonoid Metabolism in Grape Berries Grown under Double Cropping System. Front Plant Sci. 8:1912.

20. Nakaune R, Toda S, Mochizuki M, Nakano M (2008) Identification and characterization of a new vitivirus from grapevine. Arch Virol. 153:1827.

21. Alabi OJ, Poojari S, Sarver K, Martin RR, Naidu RA (2013) Complete genome sequence analysis of an American isolate of Grapevine virus E. Virus genes. 46:563-566.

22. Li H, Havens WM, Nibert ML, Ghabrial SA (2011) RNA sequence determinants of a coupled termination-reinitiation strategy for downstream open reading frame translation in Helminthosporium victoriae virus 190S and other victoriviruses (Family Totiviridae). J Virol. 85: 7343-7352.

23. Satyanarayana T, Gowda S, Ayllón MA, Albiach-Martí MR, Rabindran S, Dawson WO (2002) The p23 protein of Citrus tristeza virus controls asymmetrical RNA accumulation. J Virol. 76:473-483.

24. Vončina D & Almeida RPP (2018). Screening of some Croatian autochthonous grapevine varieties reveals a multitude of viruses, including novel ones. Arch Virol. https://doi.org/10.1007/s00705-018-3850-6

bioRxiv preprint doi: https://doi.org/10.1101/314674; this version posted August 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure and Table Legends

Figure 1. (A) Molecular characterization of grapevine virus L (GVL). Mapping pattern of virus RNA reads from library SRX3144956 to the consensus putative GVL assembly indicating coverage per position. The genome architecture of GVL is characterized by a monosegmented ssRNA(+) virus encoding for five ORFs arranged as 5'-UTR-RP-HP-MP-CP-NABP-UTR-3'. The predicted product of each ORF is depicted including the associated domains determined by NCBI-CDD. Abbreviations: RP, replicase protein; 22kDa, 22k Dalton hypothetical protein; MP, movement protein; CP, coat protein; NABP, nucleic acid binding protein. See the text for additional domain information. (**B**) Phylogenetic insights of putative grapevine virus L (GVL) in relation to accepted and proposed members of the genus *Vitivirus* based on MAFFT alignments (BLOSUM62 scoring matrix using as best-fit algorithm E-INS-i) of the predicted replicase and capsid protein (**C**) followed by maximum likelihood trees generated by FastTree (best-fit model = JTT-Jones-Taylor-Thorton with single rate of evolution for each site = CAT) computing local support values with the Shimodaira-Hasegawa test (SH) and 1,000 resamples). Accession numbers of the corresponding sequences: actinidia virus A (JN427014), actinidia virus B (NC_016404), mint virus 2 (AY913795), grapevine virus B (NC_003602), grapevine virus H (MF521889), grapevine virus F (NC_018458), grapevine virus A (NC_003604), grapevine virus K (NC_035202), arracacha virus V (NC_034264), grapevine virus E (NC_011106), agave tequilana leaf virus (NC_034833), grapevine virus D (KX828708), grapevine virus G (MF405923), grapevine virus I (MF927925), heracleum latent virus (X79270), and grapevine Pinot gris virus (*Trichovirus*, NC_015782) used as outgroup. Scale bar represents substitutions per site. Node labels are FastTree support values. Asterisk indicates GVL.

Figure 2. Phylogenetic insights integrating four isolates of grapevine virus L (GVL) and five isolates of grapevine virus E in the context of members of the genus *Vitivirus* based on MAFFT alignments (BLOSUM62 scoring matrix using as best-fit algorithm E-INS-i) of the predicted replicase (**A**) and capsid protein (**B**) followed by maximum likelihood trees generated by FastTree (best-fit model = JTT-Jones-Taylor-Thorton with single rate of evolution for each site = CAT) computing local support values with the Shimodaira-Hasegawa test (SH) and 1,000 resamples). Accession numbers of the corresponding sequences: grapevine virus E (GVE) isolate VVL-101(MF991950), GVE isolate WAHH2 (JX402759), GVE isolate SA94 (GU903012), GVE isolate GFMG-1 (KF588015), GVE isolate TvAQ7 (NC_011106). Additional sequences are indicated in Figure 1.B legend. Pinot gris virus was used as outgroup. Scale bar represents substitutions per site. Node labels are FastTree support values. GVL isolates are depicted in bold.

Table 1. Diverse structural highlights of putative grapevine virus L in the context of proposed and ICTV recognized (in italics) *Vitivirus* species. *Grapevine Pinot gris virus* (genus *Trichovirus*) is included as outgroup for phylogenetic analyses. Abbreviations: Accession #: NCBI Genbank accession number; Genome: Genome size (nt), 5´U: 5´UTR length (nt); RP: replicase protein length (aa); HP: hypothetical protein length (aa); MP: movement protein length (aa); CP: coat protein length (aa); NABP: nucleic acid binding protein length (aa); 3´UTR: 3´UTR length excluding the poly A tail (nt).

bioRxiv preprint doi: https://doi.org/10.1101/314674; this version posted August 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Figure 1. ClustalW alignment of 5'-UTR (**A**) and 3'-UTR (**B**) of GVL, GVE and the proposed GVI and GVG. GVL 5'-UTR terminal 20 nt are identical to that of GVI. At the 3'-UTR in an equidistal position of 3' terminus a ten nucleotides monomer "GGCGAAAUAA" is conserved among the four virus sequences.

Supplementary Figure 2. Genetic distances expressed as percentage values and heat-maps of predicted RP (A) and CP (B) of accepted and proposed members of the genus *Vitivirus* and GVL, based on MAFFT alignments (BLOSUM62 scoring matrix using as best-fit algorithm E-INS-I). See Table 1 for abbreviations.

Supplementary Figure 3. Phylogenetic insights of putative grapevine virus L (GVL) in the context of family *Betaflexiviridae*, based on MAFFT alignments (BLOSUM62 scoring matrix using as best-fit algorithm E-INS-i) of NCBI refseq replicase proteins of *Betaflexiviridae* members, followed by maximum likelihood radial trees generated by FastTree (best-fit model = JTT-Jones-Taylor-Thorton with single rate of evolution for each site = CAT) computing local support values with the Shimodaira-Hasegawa test (SH) and 1,000 resamples). Scale bar represents substitutions per site. *Betaflexiviridae* family is depicted in blue and the monophyletic clade of vitiviruses is shown in red. GVL is indicated with an asterisk.

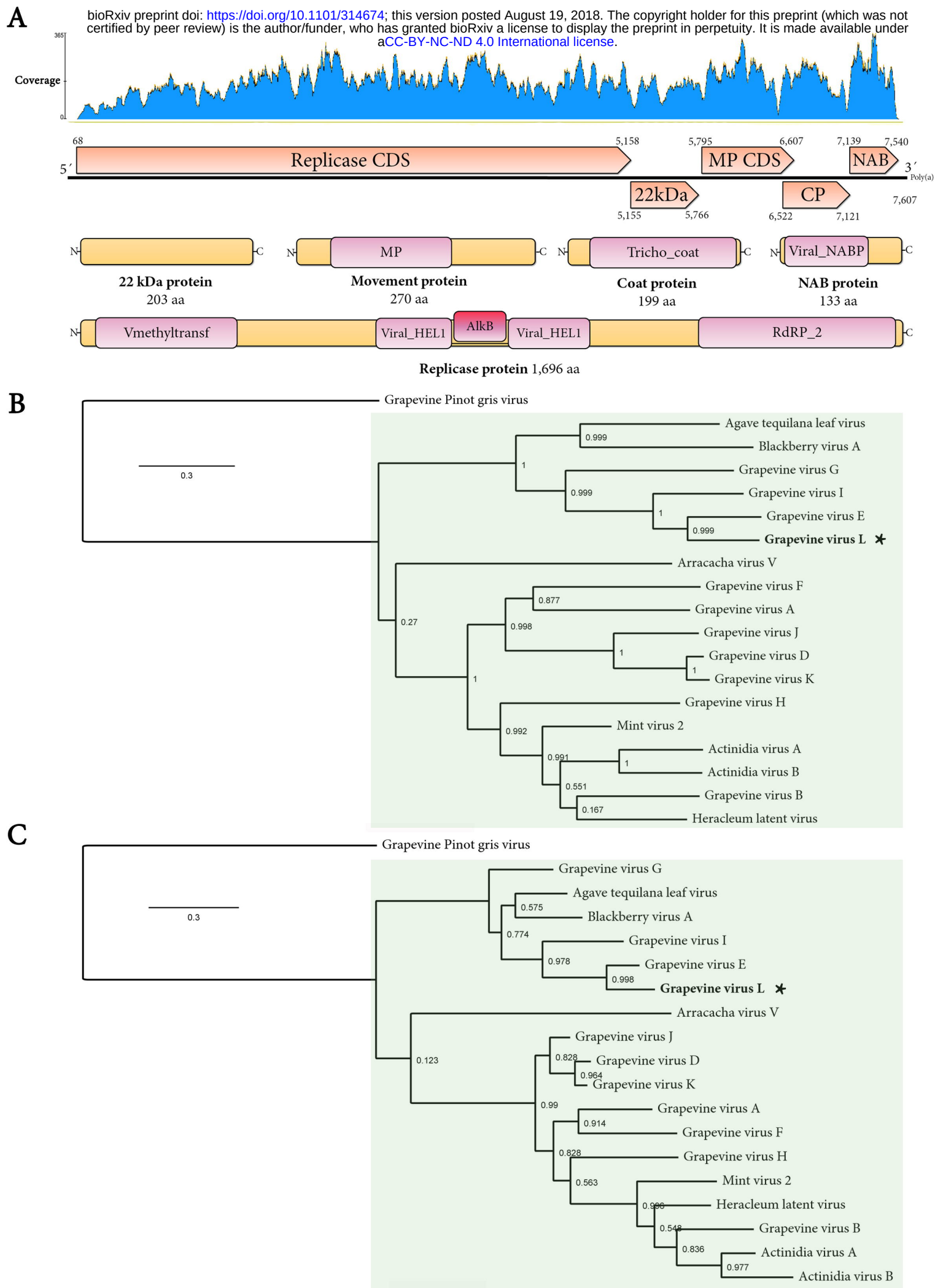
Supplementary Figure 4. Genomic architecture of the draft genomes of the identified GVL isolates characterized by a monosegmented ssRNA(+) virus encoding for five ORFs arranged as 5'-UTR-RP-HP-MP-CP-NABP-UTR-3'. The predicted product of each ORF is depicted including the associated domains determined by NCBI-CDD. Abbreviations: RP, replicase protein; 22kDa, 22k Dalton hypothetical protein; MP, movement protein; CP, coat protein; NABP, nucleic acid binding protein.

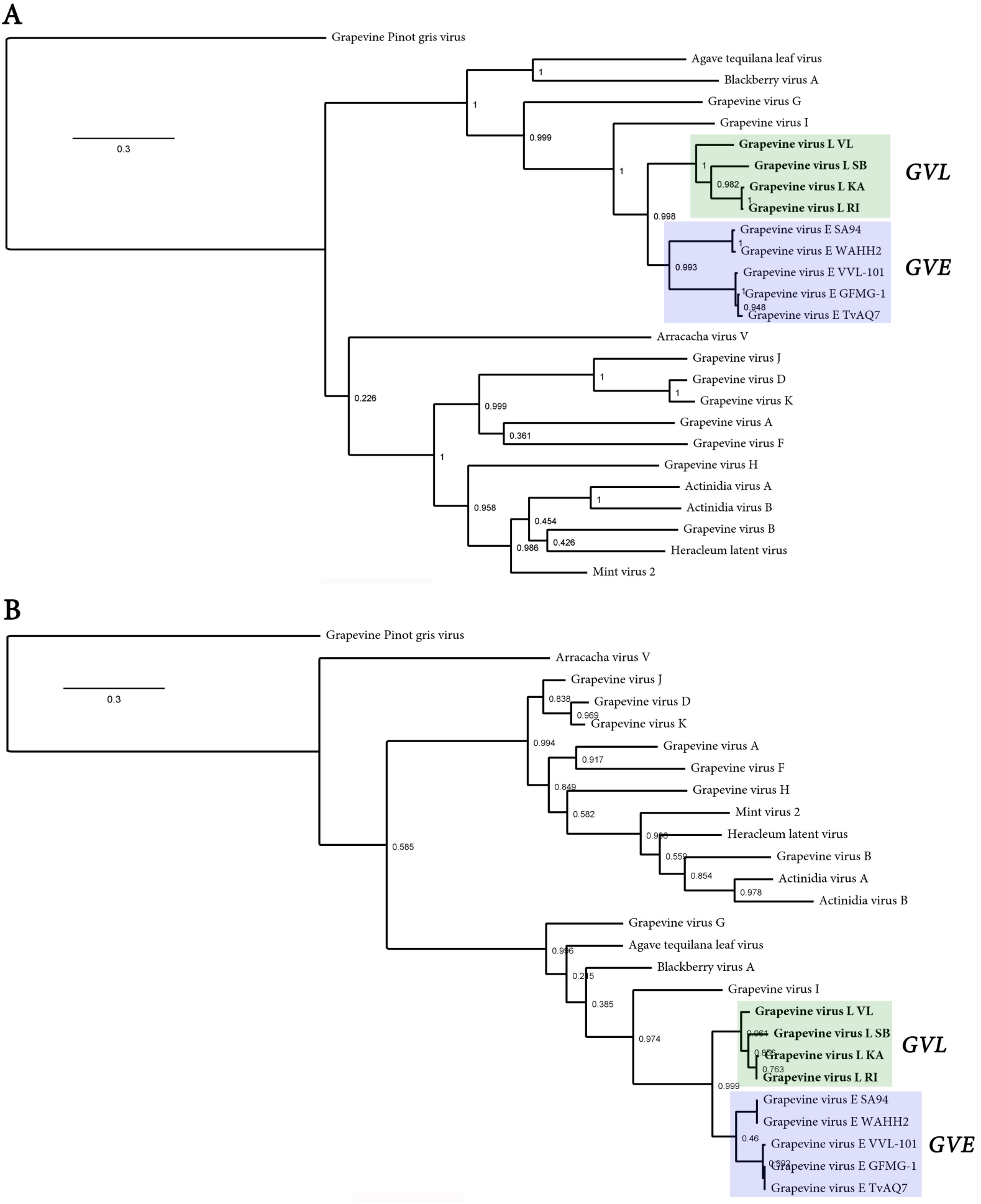
Supplementary Table 1. NGS RNA libraries used in this study generated by Chen et al [10]. Stage: sampled grapeberries developmental stage; Total reads: Total raw 50 nt reads generated for each library; GVL virus reads: viral reads detected in each sample; GVL RPM: virus reads per million total library reads.

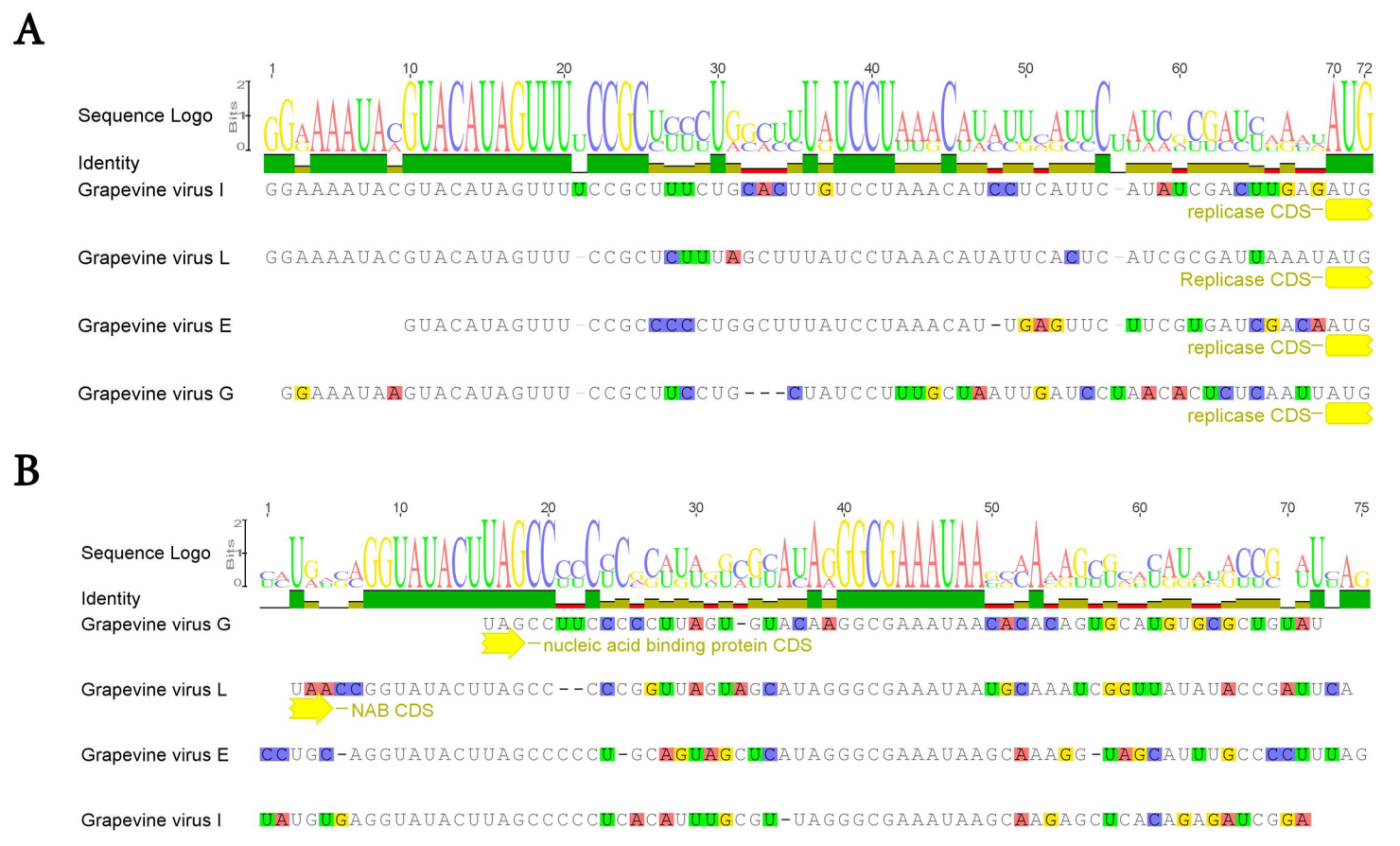
Supplementary Table 2. Variable sites of grapevine virus L identified in the virus RNA reads detected in SRX3144939- SRX3144940 (E-L 35 cv Riesling grapeberries). Polymorphism are indicated in relation with the consensus sequence of the E-L-38 cv Riesling strain determined for SRX3144956. Only variable sites with a frequency >0.5 of virus reads are shown.

Supplementary Table 3. Structural highlights and genetic distance of grapevine virus L isolates. Abbreviations: Accession #: NCBI Genbank accession number; GS: Genome size (nt), 5´U: 5´UTR length (nt); RP cds: replicase cds length (nt); HP cds: hypothetical cds length (nt); MP cds: movement protein cds length (nt); CP cds: coat protein cds length (nt); NABP: nucleic acid binding protein cds length (aa); 3´UTR: 3´UTR length excluding the poly A tail (nt). RPd-NABd: genetic distance of genomic region (nt) / predicted gene product (aa) of corresponding isolate with GVL-RI.

bioRxiv preprint doi: https://doi.org/10.1101/314674; this version posted August 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.




A
Grapevine Pinot gris virus
0.3
Agave tequilana leaf virus
Blackberry virus A
Grapevine virus G
Grapevine virus I
Grapevine virus L VL
Grapevine virus L SB
Grapevine virus L KA
Grapevine virus L RI
GVL
Grapevine virus E SA94
Grapevine virus E WAHH2
Grapevine virus E VVL-101
Grapevine virus E GFMG-1
Grapevine virus E TvAQ7
GVE
Arracacha virus V
Grapevine virus J
Grapevine virus D
Grapevine virus K
Grapevine virus A
Grapevine virus F
Grapevine virus H
Actinidia virus A
Actinidia virus B
Grapevine virus B
Heracleum latent virus
Mint virus 2
1
0.999
0.998
0.982
0.993
0.948
0.226
0.361
0.958
0.454
0.986
0.426
B
Grapevine Pinot gris virus
0.3
Arracacha virus V
Grapevine virus J
Grapevine virus D
Grapevine virus K
Grapevine virus A
Grapevine virus F
Grapevine virus H
Mint virus 2
Heracleum latent virus
Grapevine virus B
Actinidia virus A
Actinidia virus B
Grapevine virus G
Agave tequilana leaf virus
Blackberry virus A
Grapevine virus I
Grapevine virus L VL
Grapevine virus L SB
Grapevine virus L KA
Grapevine virus L RI
GVL
Grapevine virus E SA94
Grapevine virus E WAHH2
Grapevine virus E VVL-101
Grapevine virus E GFMG-1
Grapevine virus E TvAQ7
GVE
0.838
0.969
0.994
0.917
0.849
0.582
0.559
0.854
0.978
0.585
0.385
0.974
0.763
0.999
0.46

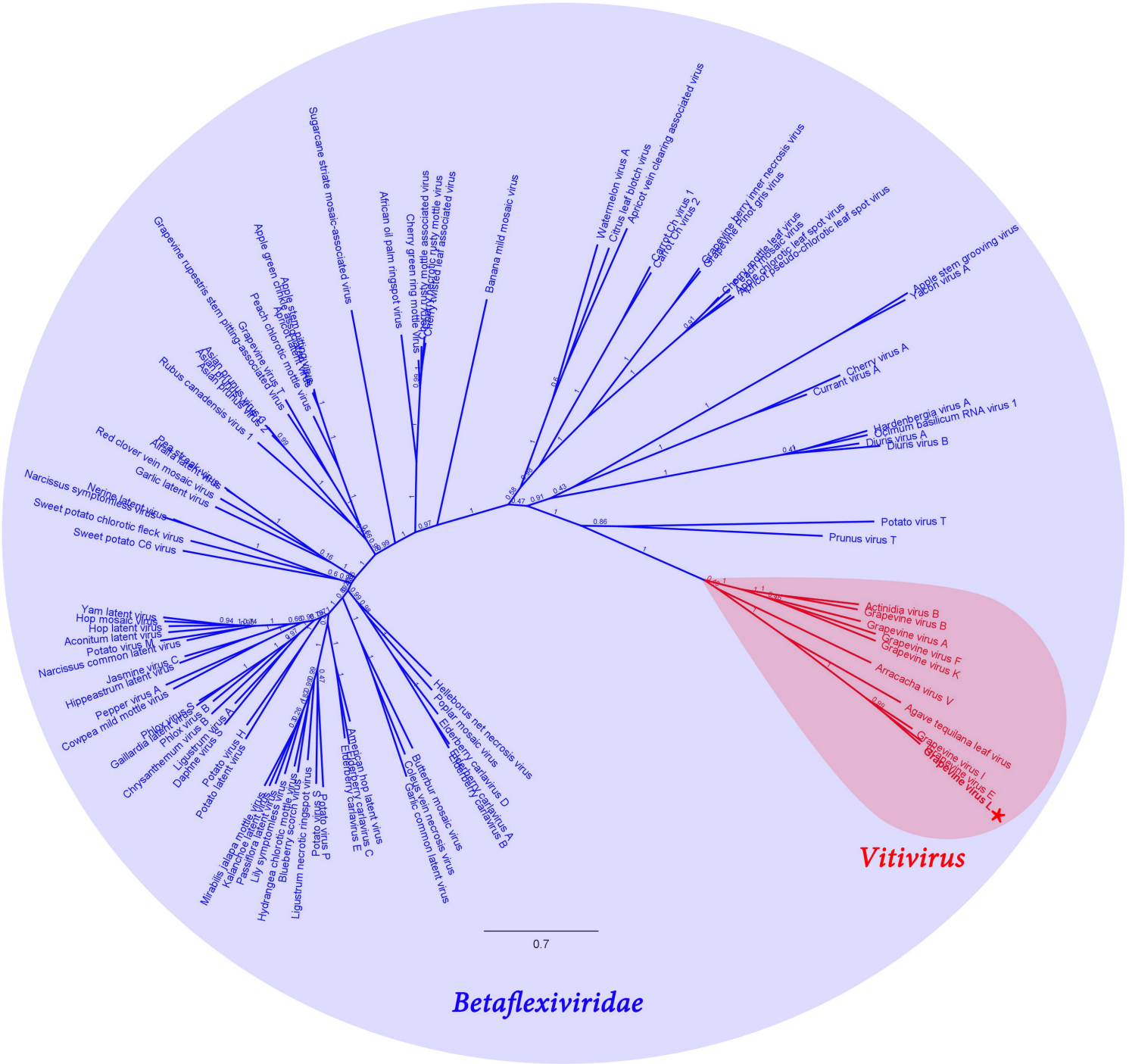

A
Sequence Logo
Identity
Grapevine virus I
GGAAAAUACGUACAUAGUUUUCCGCUUUCUGCACUUGUCCUAAACAUCCUCAUUC-AUAUCGACUUGAGAUG
replicase CDS
Grapevine virus L
GGAAAAUACGUACAUAGUUU-CCGCUCUUUAGCUUUAUCCUAAACAUAUUCACUC-AUCGCGAUUAAAUAUG
Replicase CDS
Grapevine virus E
GUACAUAGUUU-CCGCCCCCUGGCUUUAUCCUAAACAU-UGAGUUC-UUCGUGAUCGACAAUG
replicase CDS
Grapevine virus G
GGAAAUAAGUACAUAGUUU-CCGCUUCCUG---CUAUCCUUUGCUAAUUGAUCCUAACACUCUCAAUUAUG
replicase CDS
B
Sequence Logo
Identity
Grapevine virus G
UAGCCUUCCCCCUUAGU-GUACAAGGCGAAAUAACACACAGUGCAUGUGCGCUGUAU
nucleic acid binding protein CDS
Grapevine virus L
UAACCGGUAUACUUAGCC--CCCGGUUAGUAGCAUAGGGCGAAAUAAUGCAAAUCGGUUAUAUACCGAUUCA
NAB CDS
Grapevine virus E
CCUGC-AGGUAUACUUAGCCCCCU-GCAGUAGCUCAUAGGGCGAAAUAAGCAAAGG-UAGCAUUGCCCCUUUAG
Grapevine virus I
UAUGUGAGGUAUACUUAGCCCCCUCACAUUUGCGU-UAGGGCGAAAUAAGCAAGAGCUCACAGAGAUCGGA

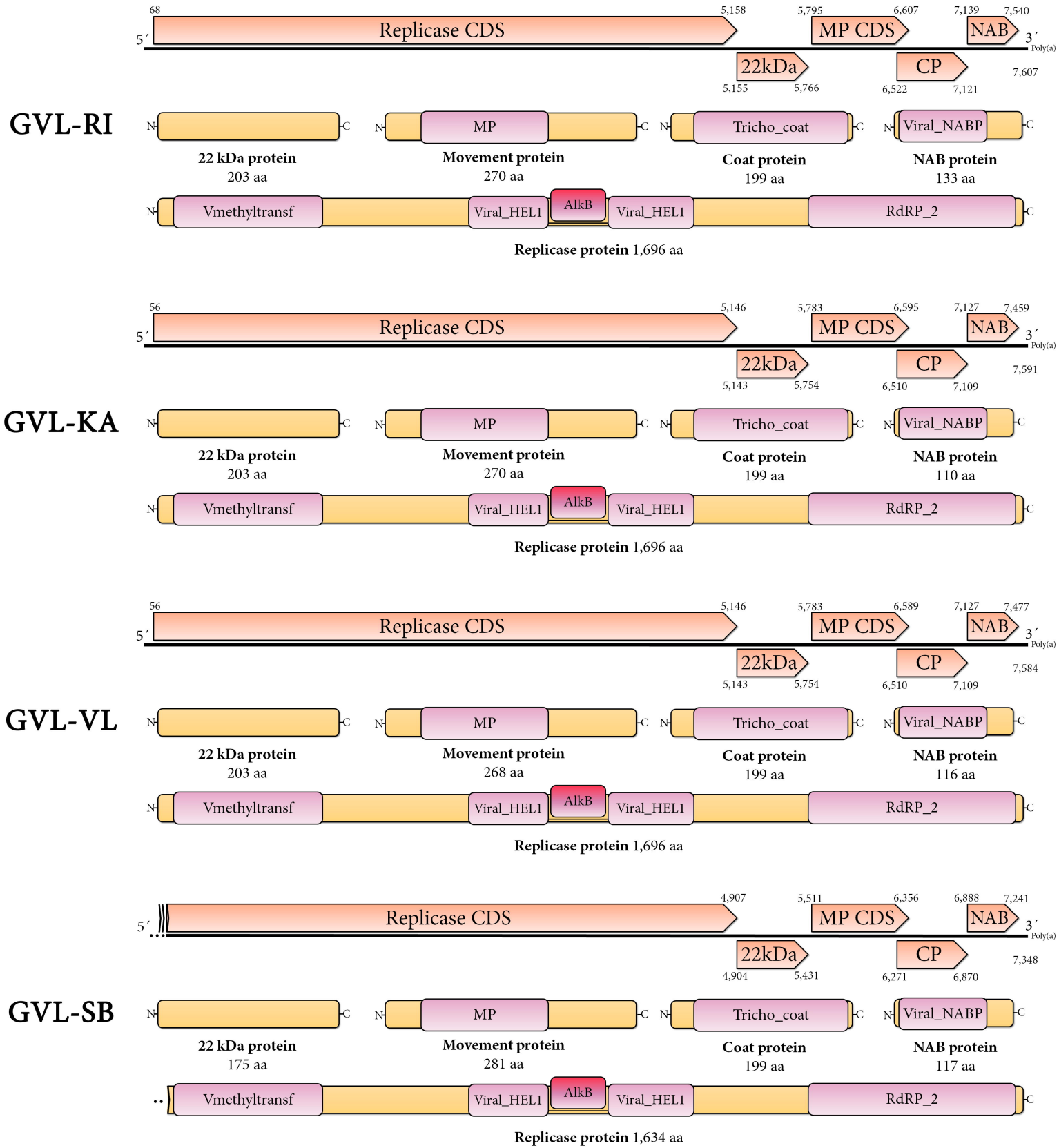
A

	HLV	MV-2	AcVB	AcVA	GVB	GVH	GVF	GVA	GVD	GVK	GVJ	GVI	* GVL	GVE	GVG	ATLV	AVV
HLV		34.6%	9.8%	9.9%	9.5%	9.1%	9.1%	8.9%	9.0%	8.9%	9.2%	8.0%	8.2%	8.2%	7.6%	8.4%	7.2%
MV-2	34.6%		72.8%	73.2%	70.6%	67.3%	66.2%	66.2%	64.4%	63.6%	66.1%	54.8%	56.3%	56.0%	54.9%	56.0%	51.9%
AcVB	9.8%	72.8%		65.9%	56.0%	50.1%	46.7%	47.1%	45.8%	45.9%	46.2%	30.5%	31.1%	31.5%	30.3%	35.1%	37.4%
AcVA	9.9%	73.2%	65.9%		54.8%	50.7%	45.5%	46.8%	45.9%	46.0%	45.5%	30.2%	31.0%	31.2%	29.0%	33.9%	38.0%
GVB	9.5%	70.6%	56.0%	54.8%		48.1%	46.7%	47.7%	46.1%	46.4%	46.6%	31.7%	31.4%	32.0%	30.6%	34.4%	37.6%
GVH	9.1%	67.3%	50.1%	50.7%	48.1%		45.8%	46.5%	45.6%	45.5%	44.9%	30.4%	30.9%	31.7%	30.3%	35.0%	37.8%
GVF	9.1%	66.2%	46.7%	45.5%	46.7%	45.8%		49.0%	48.5%	48.6%	48.6%	29.7%	30.2%	30.5%	30.8%	35.0%	36.4%
GVA	8.9%	66.2%	47.1%	46.8%	47.7%	46.5%	49.0%		50.2%	50.0%	50.1%	30.3%	30.7%	30.3%	30.7%	34.1%	37.2%
GVD	9.0%	64.4%	45.8%	45.9%	46.1%	45.6%	48.5%	50.2%		88.9%	65.6%	30.0%	30.6%	30.6%	29.3%	35.4%	37.9%
GVK	8.9%	63.6%	45.9%	46.0%	46.4%	45.5%	48.6%	50.0%	88.9%		65.1%	30.2%	30.6%	30.7%	29.1%	35.1%	37.4%
GVJ	9.2%	66.1%	46.2%	45.5%	46.6%	44.9%	48.6%	50.1%	65.6%	65.1%		29.9%	31.2%	31.0%	30.1%	35.1%	38.4%
GVI	8.0%	54.8%	30.5%	30.2%	31.7%	30.4%	29.7%	30.3%	30.0%	30.2%	29.9%		64.7%	64.2%	51.6%	41.7%	28.3%
* GVL	8.2%	56.3%	31.1%	31.0%	31.4%	30.9%	30.2%	30.7%	30.6%	30.6%	31.2%	64.7%		70.2%	51.0%	41.5%	28.8%
GVE	8.2%	56.0%	31.5%	31.2%	32.0%	31.7%	30.5%	30.3%	30.6%	30.7%	31.0%	64.2%	70.2%		50.7%	40.8%	28.9%
GVG	7.6%	54.9%	30.3%	29.0%	30.6%	30.3%	30.8%	30.7%	29.3%	29.1%	30.1%	51.6%	51.0%	50.7%		41.5%	27.3%
ATLV	8.4%	56.0%	35.1%	33.9%	34.4%	35.0%	35.0%	34.1%	35.4%	35.1%	35.1%	41.7%	41.5%	40.8%	41.5%		33.6%
AVV	7.2%	51.9%	37.4%	38.0%	37.6%	37.8%	36.4%	37.2%	37.9%	37.4%	38.4%	28.3%	28.8%	28.9%	27.3%	33.6%	

B

	AVV	GVB	AcVA	AcVB	HLV	MV-2	GVD	GVK	GVJ	GVA	GVH	GVF	ATLV	GVG	GVI	GVE	* GVL
AVV		44.6%	45.6%	43.5%	45.1%	43.0%	45.1%	44.0%	44.6%	41.2%	43.5%	39.2%	41.5%	39.5%	39.0%	38.0%	38.0%
GVB	44.6%		68.0%	60.9%	65.5%	62.4%	58.9%	57.9%	57.4%	55.1%	54.8%	54.0%	41.5%	36.6%	35.5%	37.5%	37.5%
AcVA	45.6%	68.0%		75.3%	67.5%	60.8%	58.4%	59.9%	57.9%	58.6%	56.9%	55.6%	41.0%	37.6%	37.0%	39.0%	37.5%
AcVB	43.5%	60.9%	75.3%		65.5%	57.4%	57.4%	58.4%	57.4%	54.5%	54.3%	56.6%	37.0%	34.7%	34.0%	36.0%	36.0%
HLV	45.1%	65.5%	67.5%	65.5%		66.5%	62.4%	59.4%	64.0%	55.6%	56.3%	57.6%	38.5%	38.6%	39.0%	38.5%	38.0%
MV-2	43.0%	62.4%	60.8%	57.4%	66.5%		58.4%	58.4%	59.4%	58.1%	53.8%	54.0%	38.0%	36.6%	35.5%	36.0%	36.5%
GVD	45.1%	58.9%	58.4%	57.4%	62.4%	58.4%		91.9%	84.3%	68.7%	62.9%	62.6%	43.5%	44.1%	39.5%	43.5%	43.0%
GVK	44.0%	57.9%	59.9%	58.4%	59.4%	58.4%	91.9%		84.3%	69.2%	63.5%	64.1%	43.0%	43.1%	38.5%	43.0%	42.5%
GVJ	44.6%	57.4%	57.9%	57.4%	64.0%	59.4%	84.3%	84.3%		69.2%	62.9%	66.2%	46.0%	44.6%	40.0%	44.5%	44.0%
GVA	41.2%	55.1%	58.6%	54.5%	55.6%	58.1%	68.7%	69.2%	69.2%		58.1%	61.6%	43.8%	39.4%	38.8%	41.8%	41.8%
GVH	43.5%	54.8%	56.9%	54.3%	56.3%	53.8%	62.9%	63.5%	62.9%	58.1%		58.6%	42.0%	41.1%	39.5%	39.0%	39.0%
GVF	39.2%	54.0%	55.6%	56.6%	57.6%	54.0%	62.6%	64.1%	66.2%	61.6%	58.6%		39.3%	39.9%	37.8%	38.8%	39.8%
ATLV	41.5%	41.5%	41.0%	37.0%	38.5%	38.0%	43.5%	43.0%	46.0%	43.8%	42.0%	39.3%		66.8%	65.3%	60.3%	56.8%
GVG	39.5%	36.6%	37.6%	34.7%	38.6%	36.6%	44.1%	43.1%	44.6%	39.4%	41.1%	39.9%	66.8%		61.8%	58.3%	54.8%
GVI	39.0%	35.5%	37.0%	34.0%	39.0%	35.5%	39.5%	38.5%	40.0%	38.8%	39.5%	37.8%	65.3%	61.8%		62.8%	62.3%
GVE	38.0%	37.5%	39.0%	36.0%	38.5%	36.0%	43.5%	43.0%	44.5%	41.8%	39.0%	38.8%	60.3%	58.3%	62.8%		78.4%
* GVL	38.0%	37.5%	37.5%	36.0%	38.0%	36.5%	43.0%	42.5%	44.0%	41.8%	39.0%	39.8%	56.8%	54.8%	62.3%	78.4%	


Sugarcane striate mosaic-associated virus
African oil palm ringspot virus
Cherry green ring mottle virus
Cherry rusty mottle associated virus
Cherry chlorotic rusty mottle virus
Cherry mottled leaf associated virus
Banana mild mosaic virus
Watermelon virus A
Citrus leaf blotch virus
Apricot vein clearing associated virus
Carrot Ch virus 1
Carrot Ch virus 2
Grapevine berry inner necrosis virus
Grapevine Pinot gris virus
Cherry mottle leaf virus
Apple mosaic virus
Apricot chlorotic leaf spot virus
Apricot pseudo-chlorotic leaf spot virus
Apple stem grooving virus
Yacon virus A
Cherry virus A
Currant virus A
Hardenbergia virus A
Ocimum basilicum RNA virus 1
Diuris virus A
Diuris virus B
Potato virus T
Prunus virus T
Actinidia virus B
Grapevine virus B
Grapevine virus A
Grapevine virus F
Grapevine virus K
Arracacha virus V
Agave tequilana leaf virus
Grapevine virus I
Grapevine virus E
Grapevine virus L
Vitivirus
Grapevine rupestris stem pitting-associated virus
Apple green crinkle virus
Apple stem pitting virus
Apricot latent virus
Peach chlorotic mottle virus
Grapevine virus T
Asian prunus virus 3
Asian prunus virus 2
Asian prunus virus 1
Rubus canadensis virus 1
Pea streak virus
Alfalfa latent virus
Red clover vein mosaic virus
Garlic latent virus
Narcissus symptomless virus
Nerine latent virus
Sweet potato chlorotic fleck virus
Sweet potato C6 virus
Yam latent virus
Hop mosaic virus
Hop latent virus
Aconitum latent virus
Potato virus M
Narcissus common latent virus
Jasmine virus C
Hippeastrum latent virus
Pepper virus A
Cowpea mild mottle virus
Phlox virus S
Gaillardia latent virus
Phlox virus B
Chrysanthemum virus B
Ligustrum virus A
Daphne virus S
Potato virus H
Potato latent virus
Mirabilis jalapa mottle virus
Kalanchoe latent virus
Passiflora latent virus
Lily symptomless virus
Hydrangea chlorotic virus
Blueberry scorch virus
Ligustrum necrotic ringspot virus
Potato virus S
Potato virus P
American hop latent virus
Elderberry carlavirus C
Elderberry carlavirus E
Hellebrus net necrosis virus
Poplar mosaic virus
Elderberry carlavirus D
Elderberry carlavirus A
Elderberry carlavirus B
Butterbur mosaic virus
Coleus vein necrosis virus
Garlic common latent virus
0.7
Betaflexiviridae


GVL-RI
68
5,158
5,795
6,607
7,139
7,540
5′
Replicase CDS
MP CDS
NAB
3′
Poly(a)
22kDa
CP
7,607
5,155
5,766
6,522
7,121
N
C
MP
Tricho_coat
Viral_NABP
22 kDa protein
203 aa
Movement protein
270 aa
Coat protein
199 aa
NAB protein
133 aa
Vmethyltransf
Viral_HEL1
AlkB
Viral_HEL1
RdRP_2
Replicase protein 1,696 aa
GVL-KA
56
5,146
5,783
6,595
7,127
7,459
5′
Replicase CDS
MP CDS
NAB
3′
Poly(a)
22kDa
CP
7,591
5,143
5,754
6,510
7,109
N
C
MP
Tricho_coat
Viral_NABP
22 kDa protein
203 aa
Movement protein
270 aa
Coat protein
199 aa
NAB protein
110 aa
Vmethyltransf
Viral_HEL1
AlkB
Viral_HEL1
RdRP_2
Replicase protein 1,696 aa
GVL-VL
56
5,146
5,783
6,589
7,127
7,477
5′
Replicase CDS
MP CDS
NAB
3′
Poly(a)
22kDa
CP
7,584
5,143
5,754
6,510
7,109
N
C
MP
Tricho_coat
Viral_NABP
22 kDa protein
203 aa
Movement protein
268 aa
Coat protein
199 aa
NAB protein
116 aa
Vmethyltransf
Viral_HEL1
AlkB
Viral_HEL1
RdRP_2
Replicase protein 1,696 aa
GVL-SB
4,907
5,511
6,356
6,888
7,241
5′
Replicase CDS
MP CDS
NAB
3′
Poly(a)
22kDa
CP
7,348
4,904
5,431
6,271
6,870
N
C
MP
Tricho_coat
Viral_NABP
22 kDa protein
175 aa
Movement protein
281 aa
Coat protein
199 aa
NAB protein
117 aa
Vmethyltransf
Viral_HEL1
AlkB
Viral_HEL1
RdRP_2
Replicase protein 1,634 aa

Virus	*Abbreviation*	*Accession #*	*Genome*	*5'UTR*	*RP*	*HP*	*MP*	*CP*	*NABP*	*3'UTR*
Actinidia virus A	AcVA	JN427014	7566	NA	1714	219	297	198	105	137
Actinidia virus B	AcVB	NC_016404	7488	70	1707	231	290	198	106	83
Agave tequilana leaf virus	ATLV	NC_034833	6958	52	1539	151	267	199	126	86
Arracacha virus V	AVV	NC_034264	7398	51	1705	NA	293	192	110	73
Blackberry virus A	BVA	MG254193	7285	59	1676	148	272	199	109	18
Grapevine virus A	GVA	NC_003604	7351	86	1707	177	278	198	90	68
Grapevine virus B	GVB	NC_003602	7599	48	1707	179	322	197	123	148
Grapevine virus D	GVD	MF774336	7479	79	1698	161	273	197	91	87
Grapevine virus E	GVE	NC_011106	7564	57	1698	191	265	199	108	134
Grapevine virus F	GVF	NC_018458	7551	90	1727	159	271	198	109	82
Grapevine virus G	GVG	MF405923	7496	64	1703	154	286	201	117	53
Grapevine virus H	GVH	MF521889	7446	97	1717	162	265	197	106	82
Grapevine virus I	GVI	NC_037058	7507	68	1696	167	264	199	121	104
Grapevine virus J	GVJ	MG637048	7390	95	1702	164	273	197	103	69
Grapevine virus K	GVK	NC_035202	7476	80	1700	161	273	197	91	77
Heracleum latent virus	HLV	X79270	3006	NA	253	NA	343	197	115	56
Mint virus 2	MV-2	AY913795	3897	NA	500	209	329	209	111	85
Grapevine virus L	GVL	MH248020	7607	67	1696	203	270	199	133	67
Vitivirus x□	NA	NA	7438	72	1692	177	286	198	108	89
Grapevine Pinot gris virus	GPGV	NC_015782	7275	94	1845	NA	369	195	NA	99

Table 1 Diverse structural highlights of grapevine virus L in the context of proposed and ICTV recognized (in italics) *Vitivirus* species. *Grapevine Pinot gris virus* (genus *Trichovirus*) is included as outgroup for phylogenetic analyses. Accession #: NCBI Genbank accession number; Genome: Genome size (nt), 5´U: 5´UTR length (nt); RP: replicase protein length (aa); HP: hypothetical protein length (aa); MP: movement protein length (aa); CP: coat protein length (aa); NABP: nucleic acid binding protein length (aa); 3´UTR: 3´UTR length excluding the poly A tail (nt).

bioRxiv preprint doi: https://doi.org/10.1101/314674; this version posted August 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.