Genomic characterization of grapevine viruses N and O: Novel vitiviruses from South Africa

David A. Read [1,2*], Genevieve D. Thompson [1,3], Nathan Le Cordeur [4], Dirk Swanevelder [1] and Gerhard Pietersen [2]

[1] Agricultural Research Council (ARC) - Biotechnology Platform, 100 Old Soutpan Road, Onderstepoort, Pretoria, 0110, South Africa

[2] Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria, South Africa

[3] Gene Vantage, 34 Monte Carlo Crescent, Kyalami Business Park, Johannesburg, 1684, South Africa

[4] Patho Solutions, Olifantskop Road, Wellington, 7655, South Africa

* - Corresponding author

Email: david.read@fabi.up.ac.za

ORCID – David A. Read

0000-0002-6614-9946

Abstract

A survey was performed on a *Vitis* cultivar collection in Stellenbosch, South Africa. Metaviromes were generated for each cultivar, using an RNAtag-seq workflow. Analysis of assembled contigs indicated the presence of two putatively novel members of the genus *Vitivirus*, provisionally named grapevine virus N (GVN) and grapevine virus O (GVO). Comparisons of amino acid identities showed that GVN and GVO are most closely related to grapevine virus G and grapevine virus E respectively. The incidence of these novel viruses within the sampling site was low, with GVO and GVN associated with only five and two cultivars respectively, of the 229 sampled.

Vitis vinifera L. (grapevine) is a significant crop within the South African agricultural sector, with the country being in the top ten for both wine and table grape production [1]. Vegetative propagation and global trade of cultivars has led to grapevine being disproportionately affected by more viruses than any other crop [2]. The exponential increase in the use of metaviromics has accelerated the discovery of grapevine pathogens, with the number of currently known

viruses approaching ninety [3]. South Africa has a long history of viticulture and consequently, a diverse set of viruses circulating within the industry.

Members of the genus *Vitivirus* (family *Betaflexiviridae*) typically have genomes of approximately 7,500 nt, with five open reading frames coding for the replication associated proteins (RAP), a "22-kDa" protein associated with vector transmission, a movement protein (MP), a coat protein (CP), and a nucleic acid binding protein (NABP) that functions as suppressor of RNA silencing [4]. Grapevine vitiviruses are common components of viral populations in South African vineyards [5], with grapevine virus A (GVA), grapevine virus B, and grapevine virus E (GVE) having been previously reported [6, 7, 8]. In general, single infections with grapevine vitiviruses are associated with very mild symptoms [9], however synergistic co-infections with other viruses, result in several severe and economically important disease phenotypes [10]. GVA and GVB are part of a complex of viruses associated with rugose wood [11] and in South Africa, are associated with Shiraz decline and corky bark disease respectively [7, 12]. In recent years, the application of metaviromics approaches have led to the discoveries of a number of new grapevine vitiviruses, such as grapevine virus F [11], grapevine virus G (GVG) [13], grapevine virus H [14] and grapevine virus L [15].

In December 2019, a total of 229 samples were collected from the *Vitis* cultivar collection at the Agricultural Research Council's (ARC) Nietvoorbij campus, as described previously [16] (GPS co-ordinates: -33.912053, 18.862291). Total RNA was isolated according to White *et al*, [17] and RNAtag-seq libraries were prepared as described [18], which were sequenced using an Illumina HiSeq 2500 instrument (Illumina, San Diego, CA, United States; ARC - Biotechnology Platform, Onderstepoort, Pretoria, South Africa), as paired-ends (2 x 125 nt) using TruSeq V4 chemistry (Illumina, San Diego, CA, United States).

Datasets were demultiplexed with Je software [19]. Trimming was performed with CLC Genomics Workbench 9 (Qiagen Bioinformatics, Aarhus, Denmark), using the following parameter values: Minimum read length of 20 nt, quality limit of 0.05 and adapter trimming with Illumina universal (5'-AGATCGGAAGAG-3') and RNAtag-seq (5'-TACACGACGCTCTTCCGATCTNNNNNNNNT-3') adapters. Assembly of trimmed reads into contigs was done with SPAdes 3.14.0 [20], with the *meta* option. The cultivars and datasets associated with this study, are listed in Supplementary Table 1. These datasets are associated with NCBI BioProject PRJNA763365. Contigs of putative viral origin, were identified using BLASTn [21] and the viral fraction of the NCBI Refseq database. Plant virus contigs were then submitted to the browser-based version of BLASTx, using the NCBI nr database, for further confirmation of their identity.

A novel vitivirus, provisionally named grapevine virus N (GVN), was associated with *V. vinifera* cultivars Azal tinto, Bourboulenc, Cinsault noir, Crouchen blanc and Palamino, while the provisionally named grapevine virus O (GVO) was associated with Roter Zierfandler and Royal Molenberg. The GVN and GVO genome organizations and associated amino acid sequences of each gene were determined using ORF finder [22], showing the typical arrangements of the genus *Vitivirus* [11]. The average amino acid identity (AAI) between each gene product and that of the closest relative was determined using the AAI calculator from the enveomics collection [23] and shown in Table 1. The amino acid sequences of ORF1 (RAP) of both GVN and GVO were aligned against cognate NCBI RefSeq sequences of extant vitiviruses, using BioEdit 7.2.5 [24]. A best-fit maximum likelihood phylogeny, based on the Le Gascuel model + G + I + F, was generated using MEGA X [25] and shown in Figure 1.

Partial segments of the GVN and GVO CP gene were amplified using RT-PCR, for each respective sample. Primers with the following sequences were used to set up two-step RT-PCR reactions: GVN-CP-F (5'-TCGCTGAGATAATAAGGAGGATTGAG-3); GVN-CP-R (5'-GACTTGAATCACACTGGCTTCAGA-3'); GVO-CP-F (5'-GGTGTGATAGAGGAT AACCACAGT-3') and GVO-CP-R (5'-TACACTCTAAACGACCACAACAGT-3'). Two-step RT-PCR reactions were carried using Promega GoScript™ Reverse Transcriptase and GoTaq® Taq polymerase (Promega, Madison, WI, United States), according to manufacturer instructions. Amplicons of the expected size were visualized on an agarose gel (572 and 789bp for GVO and GVN respectively) and the identities confirmed using bidirectional Sanger sequencing (Inqaba Biotechnical Industries, Pretoria, South Africa).

The 5' terminal nucleotides of GVN and GVO were confirmed using the 5' RACE System for Rapid Amplification of cDNA Ends, version 2.0 (Invitrogen, Carlsbad, CA, United States) according the to manufacturer's specifications. Total RNA from Palamino and Roter Zierflander was used as the inputs for GVN and GVO 5' RACE reactions respectively. The following gene specific primers were used for GVN (GVN-GSP1: 5'-CACTATATCTTAACTCATCT-3' and GVN-GSP2: 5'-CCTCTACAATATGACTAGATATGCT-3') and for GVO (GVO-GSP1: 5'-AGGGTCGTCTTATCTTCATC-3' and GVO-GSP2: 5'-TCCCTATACCTTAGGTTATC CTTAGC-3'). 3' RACE amplification was not performed as all of the GVN and GVO contigs were associated with a poly-A-tract between 48 and 52 nt in length (data not shown). Confirmation of the 5' ends showed that the GVN and GVO genomes are 7,486 and 7,560 nt respectively (excluding poly-A-tracts).

Considering the RAP phylogeny, GVN and GVO are most closely related to GVG and GVE respectively, which is also confirmed by AAI shared between the various genes of these viruses. Phylogenetically, GVN and GVO both cluster within the recently ascribed GVE superclade [26]. AAI indicates that GVN and GVO are novel viruses within the *Betaflexiviridae* family where the species demarcation limit is less than 80% AAI for either the RAP or CP genes [27]. However, some additional criteria may need to be implemented in order to determine whether GVN and GVO are representatives of two new species or distinct members of already established taxa. GVN and GVO showed limited incidence and genetic diversity within the virus population of the vineyard under study, which could suggest a recent introduction. However, more widespread surveys of South Africa's viticultural regions is needed in order to get a complete view on their distribution nationwide. Finally, the role of GVO and GVN in any grapevine disease requires further investigation, as both viruses were found in co-infections with other viruses including grapevine leafroll-associated virus 3, grapevine leafroll-associated virus 2 and at least two additional grapevine vitiviruses (data not shown). Given the potential for grapevine vitiviruses to be associated with synergistic co-infections [9], GVN and GVO should be considered possibly damaging pathogens for South African viticulture.

Declarations

Funding: David Read is grateful for the financial support provided by the National Research Foundation of South Africa, under grant UID 104901.

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

Availability of data and material: The data that support the findings of this study, are openly available in NCBI public databases.

Code availability: Not applicable

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

References

1 Gbejewoh O, Keesstra S, Blancquaert E (2021) The 3Ps (Profit, Planet, and People) of Sustainability amidst Climate Change: A South African Grape and Wine Perspective. Sustainability 13:2910

2 Martelli GP (2018) Where grapevine virology is heading to. In: Proceedings of the 19th congress of the International Council for the Study of virus and virus-like diseases of the grapevine (ICVG), Santiago, Chile, 9-12 April 2018

3 Fuchs M (2020) Grapevine viruses: a multitude of diverse species with simple but overall poorly adopted management solutions in the vineyard. J Plant Pathol 102:643–653

4 Galiakparov N, Tanne E, Sela I, Gafny R (2003) Functional analysis of the grapevine virus a genome. Virology 306:42–50.

5 du Preez J, Stephan D, Mawassi M, Burger JT (2011) The grapevine-infecting vitiviruses, with particular reference to grapevine virus A. Arch Virol 156:1495–1503

6 Engelbrecht DJ, Kasdorf GGF (1987) Occurrence and transmission of grapevine virus A in South African grapevines. South African J Enol Vitic 8(1):23-29.

7 Goszczynski DE (2010) Divergent molecular variants of Grapevine virus B (GVB) from corky bark (CB)-affected and CB-negative LN33 hybrid grapevines. Virus Genes 41:273–281

8 Goszczynski DE, Jooste AEC (2003) Identification of divergent variants of Grapevine virus A. Eur J Plant Pathol 109:397–403

9 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 151: 919–925

10 Rowhani A, Uyemoto JK, Golino D, Daubert SD, Al Rwahnih M (2016) Viruses involved in graft- incompatibility and decline. In Meng B, Fuchs M, Martelli G Golino D (eds) Grapevine viruses: Molecular biology, diagnostics, and management, Chapter 12. Springer, New York, pp 289–302

11 Al Rwahnih M, Sudarshana MR, Uyemoto JK, Rowhani A (2012) Complete genome sequence of a novel vitivirus isolated from grapevine. J Virol 86(17):9545

12 Goszczynski DE, du Preez J, Burger JT (2008) Molecular divergence of Grapevine virus A (GVA) variants associated with Shiraz disease in South Africa. Virus Res 138:105–110

13 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

14 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

15 Debat HJ, Zavallo D, Brisbane RS, Voncina D, Almeida RPP, Blouin AG, Al Rwahnih M, Gomez TG, Asurmendi S (2019) Grapevine virus L: A novel vitivirus in grapevine. Eur J Plant Pathol 155:319–328.

16 Read DA, Thompson GD, Swanevelder D, Pietersen G (2021) Detection and diversity of grapevine virus L from a *Vitis* cultivar collection in Stellenbosch, South Africa. Eur J Plant Pathol https://doi.org/10.1007/s10658-021-02380-y

17 White EJ, Venter M, Hiten NF, Burger JT (2008) Modified cetyltrimethylammonium bromide method improves robustness and versatility: the benchmark for plant RNA extraction. Biotechnol J 3:1424–1428

18 Shishkin AA, Giannoukos G, Kucukural A et al (2015) Simultaneous generation of many RNA-seq libraries in a single reaction. Nat Meth 12:323–325

19 Girardot C, Scholtalbers J, Sauer S et al (2016) Je, a versatile suite to handle multiplexed NGS libraries with unique molecular identifiers. BMC Bioinformatics 17(1):419

20 Nurk S, Meleshko D, Korobeynikov A, Pevzner PA (2017) metaSPAdes: a new versatile metagenomic assembler. Genome Res 27(5):824-834

21 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

22 Wheeler DL, Church DM, Federhen S, Lash AE, Madden TL, Pontius JU, Schuler GD, Schriml LM, Sequeira E, Tatusova TA, Wagner L (2003) Database resources of the National Center for Biotechnology. Nucleic Acids Res 31:28-33

23 Rodriguez-R LM, Konstantinidis KT (2016) The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ Prepr 4:e1900v1

24 Hall TA (1999) BioEdit: A user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. Nucleic Acids Symp Ser 41:95-98

25 Stecher G, Tamura K, Kumar S (2020) Molecular Evolutionary Genetics Analysis (MEGA) for macOS. Mol Biol Evol 37:1237–1239

26 Maree HJ, Blouin AG, Diaz- Lara A, Mostert I, Al Rwahnih M, Candresse T (2020) Status of the current vitivirus taxonomy. Arch Virol 165:451–458

27 Adams MJ, Antoniw JF, Bar-Joseph M, Brunt AA, Candresse T, Foster GD, Martelli GP, Milne RG, Zavriev SK, Fauquet CM (2004) The new plant virus family Flexiviridae and assessment of molecular criteria for species demarcation. Arch Virol 149:1045–1060

Table 1: Data on grapevine virus N (GVN) and grapevine virus O (GVO) genomes and deduced genome products. AAI - amino acid sequence identity shared between the viruses in this study and cognate gene products the most closely related viruses MW – molecular weight in kilodaltons; pI – isoelectric point; RAP – replicase-associated proteins; ORF2 – open reading frame 2; MP – movement protein; CP – coat protein; NABP – nucleic acid binding protein. [1] NC040616 grapevine virus G [2] NC011106 grapevine virus E

Virus	Cultivar/ GenBank acc.	Length (nt)	Gene	Gene location (nt)	Product size (aa)	Putative gene function/s	MW (kDa)	pI	AAI
GVN	25-01 Palomino/ MZ682355	7,486	RAP	64-5178	1704	Replication	194	5.9	70.6[1]
			ORF2	5184-5660	158	Hypothetical protein	17.7	6.5	36.8[1]
			MP	5681-6529	282	Movement protein	31.6	5.8	67.3[1]
			CP	6441-7046	201	Coat protein	21.8	7.8	80.1[1]
			NABP	7081-7413	110	RNA-binding protein	12.6	7.4	87.0[1]
GVO	28-12 Roter Zierfandler/ MZ682356	7,560	RAP	68-5161	1697	Replication	192	6.1	75.6[2]
			ORF2	5158-5724	188	Hypothetical protein	21.2	6.8	53.3[2]
			MP	5752-6558	268	Movement protein	29.6	6.0	77.7[2]
			CP	6434-7078	214	Coat protein	23.6	9.0	87.4[2]
			NABP	7096-7425	109	RNA-binding protein	12.6	9.6	69.1[2]

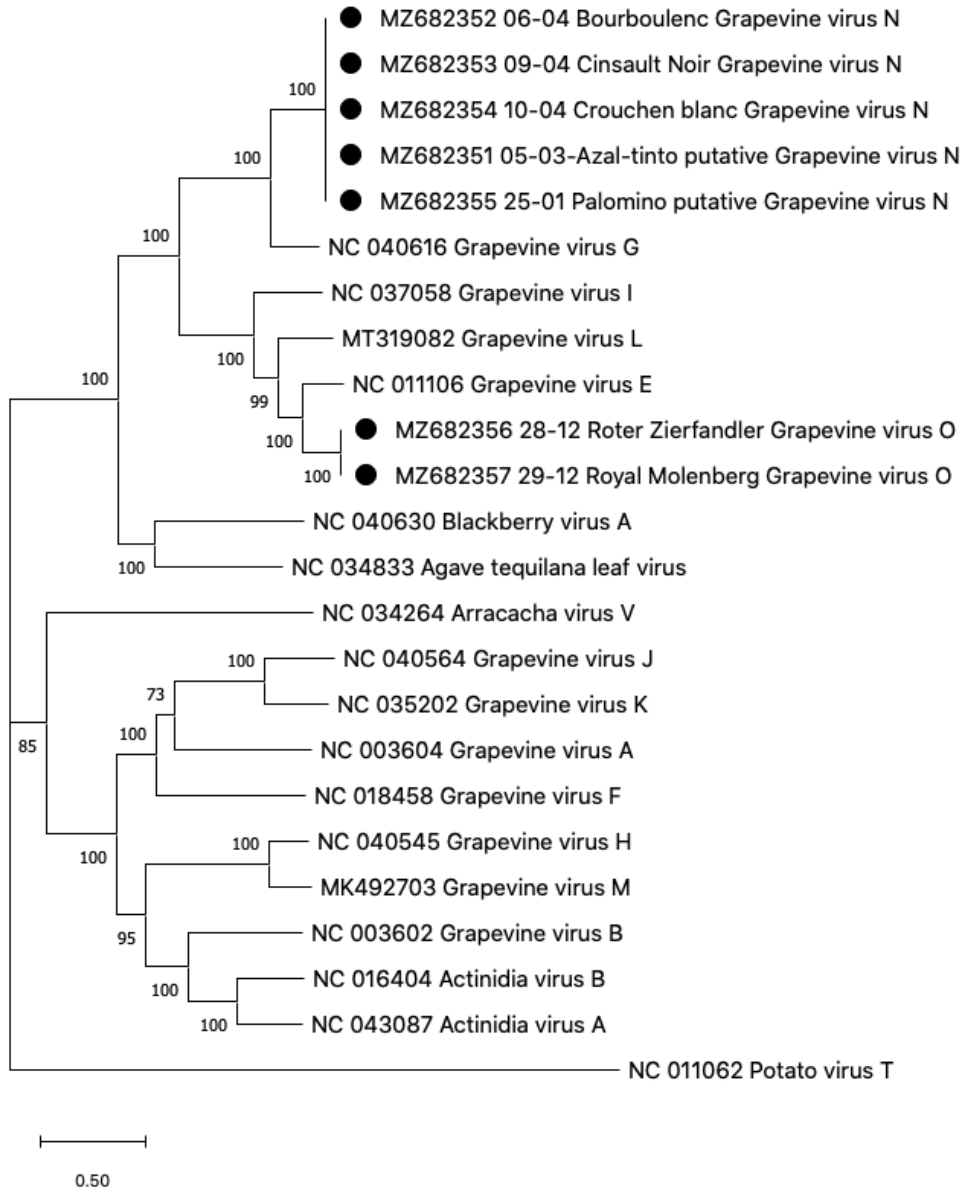

Figure 1: Maximum likelihood phylogeny based on the amino acid sequences derived from ORF 1 (replicase associated proteins; RAP) of grapevine virus N and grapevine virus O (indicated by solid circle markers) from this study and references derived from other extant vitiviruses. The cognate RAP sequence from potato virus T was used as an outgroup. The phylogeny represents the tree with the highest log likelihood and was generated in MEGA X using the Le Gascuel model with frequencies, invariant sites and gamma distribution (n=5). Bootstrapping was applied (1000 replicates) and the percentage of trees in which the associated taxa clustered together is shown next to the branches. Bootstrap percentages lower than 50 are not shown.