CrossMark

ORIGINAL ARTICLE

High-throughput sequencing applied for the identification of viruses infecting grapevines in Brazil and genetic variability analysis

Thor V. M. Fajardo[1] · Fábio N. Silva[2] · Marcelo Eiras[3] · Osmar Nickel[1]

Abstract The application of high-throughput sequencing technologies (HTS) enables the recovery of many nucleotide sequence fragments from diseased plants and may help in pathogen identification. This study was designed to identify viruses infecting 15 grapevine (*Vitis* spp.) samples collected from experimental fields and vine collections and assess the genetic variability of the identified viruses. The virus-enriched dsRNAs were extracted from bark scrapings and sequenced using an Illumina platform. The paired-end reads were analyzed, assembled contigs were generated and identified as related to viruses. Contigs of 14 viruses have been identified, some of them covering large extensions of viral genomes or resulting in assembly of near-complete or complete genomes. Grapevine virus infections are usually mixed and the HTS assays were suitable to identify ten viruses already reported that traditionally infect grapevines in Brazil, one that has been recently identified (*Grapevine Syrah virus 1*) and others (Grapevine Cabernet Sauvignon reovirus, *Grapevine Red Globe virus* and *Grapevine vein clearing virus*) not previously reported in this country. Nucleotide identities among Brazilian isolates identified by HTS and homologous grapevine virus

sequences in GenBank were high, ranging from 77% to 99%. Genetic variability analysis of viral sequences obtained by HTS and sequences available in GenBank indicated that the coding regions in the different viral species are under purifying selection, and that recombination events occurred in the majority of the viral species analyzed. The coat protein genes, generally, had lower genetic variability than the replicase and movement protein genes.

Keywords *Vitis* · Diagnosis · HTS · Next-generation sequencing · NGS · Variability

Introduction

Grapevine (*Vitis* spp.) is one of the most important fruit crops worldwide. Among grapevine pathogens, viruses stand out, considering that this crop is susceptible to numerous graft-transmissible agents that cause several diseases negatively affecting plant vigor, production and fruit quality. To date, nearly 70 virus species have been identified that are able to infect plants from the *Vitis* genus. Most of them are viruses with RNA genomes, however, at least five viruses with DNA genomes have also been reported from grapevines recently. These viruses are identified and classified based on parameters such as particle size, genome structure (ORFs), nucleotide or amino acid sequence identity of different virus proteins, type of transmission by vectors and serological information. It should be noted that, in many cases, they are found associated as a multiple virus infection complex (Martelli 2014; Naidu et al. 2015; Armijo et al. 2016).

The application of accurate and reliable diagnosis methods has fundamental importance in programs that are intended to produce and maintain virus-free propagating materials, and are very important to the success of any preventive and control

Section Editor: Michael Goodin

✉ Thor V. M. Fajardo
thor.fajardo@embrapa.br

[1] Embrapa Uva e Vinho, Bento Gonçalves, RS 95701-008, Brazil

[2] Universidade do Estado de Santa Catarina, Lages, SC 88520-000, Brazil

[3] Instituto Biológico de São Paulo, São Paulo, SP 04014-002, Brazil

strategies of grapevine viral diseases (Maliogka et al. 2015). Among standard and traditional detection methods for identification of grapevine viruses are the screening methods for a range of suspected known viruses using a panel of specific tests, such as ELISA, RT-PCR and RT-qPCR. Their disadvantages are that serological or molecular detection methods for plant viruses require availability of antibody or prior knowledge of nucleotide sequence, so they are limited in their ability of detecting variants, such as new strains or unknown agents as well as previously non-defined target viruses. Thus, it is necessary to move to broader techniques when negative results are obtained from samples suspected of virus infection (Basso et al. 2017).

In contrast, high-throughput sequencing (HTS), also known as next-generation sequencing (NGS), appears as a powerful technology that allows detection of previously non-defined target viruses and discovery of viruses without any prior knowledge of their genomes (Massart et al. 2014; Burger and Maree 2015; Roossinck et al. 2015; Hadidi et al. 2016). The application of HTS technologies enables the recovery of thousands of sequence fragments from diseased plants and may help in the identification of mixed and unknown pathogens. Also, the capacity to analyze asymptomatic viral infections or coinfections showed by HTS technologies is especially suitable to apply in grapevines because this situation often occurs with this host. Several metagenomic studies have been developed aiming at characterizing viromes (including causal agents of economically important grapevine diseases), leading to the identification of new viral species or strains of known viruses from RNA, siRNA or dsRNA isolated from grapevines (Al Rwahnih et al. 2009, 2013, 2015; Coetzee et al. 2010; Giampetruzzi et al. 2012; Jo et al. 2015; Poojari et al. 2013; Zhang et al. 2011).

Due to the continued recognition of the threat of viruses to the sustainability of the grapevine industry, considerable investments are being made to improve the capacity to accurately and rapidly identify and characterize viruses with potential to cause economical damages. Additionally, efforts are made to control their impact before they become widespread, and HTS technology has the potential to enhance these control efforts. The comprehensive nature of HTS analysis, its analytical speed and greater sensitivity over traditional methods improve surveys of recognized viruses and the discovery of novel ones (Al Rwahnih et al. 2015).

Sequences generated by HTS can also be used in genetic variability studies. There are few studies of diversity and genetic variability of Brazilian isolates of grape viruses. Determining which species are prevalent in specific regions, as well as performing genetic variability studies are important steps to understand how pathogens evolve and change, predicting epidemics and recommending control measures (Moura et al. 2017). Based on genetic variability studies it is also possible to develop more consistent tools (such as specific primers) for diagnosis of plant viruses, by revealing more conserved regions in the genome of a particular viruses. The aims of this work were to identify viruses infecting grapevine samples collected from experimental fields and vine collections, and to perform molecular characterization and genetic variability studies of the identified viruses.

Material and methods

Plant material Fifteen grapevine (*Vitis* spp.) samples collected from experimental fields, germplasm, genotype and virus collections maintained by Brazilian research centers were chosen to identify infecting viruses and assess the genetic variability of the identified viruses. The analysed plants were symptomatic (i.e. down rolling and reddening/yellowing of leaves, coriaceous leaves), or asymptomatic, and included grape cultivars and wild grapevines (Table 1).

HTS Double-stranded RNAs (dsRNA) were extracted from 30 g of bark scrapings of mature grapevine canes per sample. For dsRNA analyses, nucleic acids were first purified through a phenol:chloroform extraction. Briefly, plant tissue was ground in liquid nitrogen and 90 mL STE 2x (0.2 M NaCl, 0.1 M Tris, 0.002 M EDTA, pH 7.5), 35 mL SDS (10% w/v), 2 mL bentonite (45 mg/mL) and 2 mL β-mercaptoethanol were added. After shaking for 5 min at room temperature, 35 mL of buffer-saturated phenol, pH 7.5 and 35 mL of chloroform:isoamyl alcohol (24:1 v/v) were added, followed by shaking for 45 min and centrifugation at 16,000 g for 10 min at 4 °C. The resulting supernatant (nucleic acid solution) was enriched for dsRNA using Whatman CF11 cellulose affinity chromatography (Valverde et al. 1990). Final eluates were cleaned up using a commercial kit and checked regarding quality and concentration of dsRNA by spectrophotometry (NanoDrop 2000, Thermo Scientific). Sequencing data were generated from a complementary DNA library, which was constructed by Macrogen Inc. (Seoul, South Korea) or Eurofins Genomics Company (Huntsville, USA). The Illumina HiSeq 2000 platform was used to generate the paired-end reads. CLC Genomics Workbench software v.6.0.3 (CLC Bio, Qiagen) or Geneious 7.1.7 software (Biomatters Ltd.) were used for quality trimming and *de novo* contig assembly from the reads. All contigs were analyzed using NCBI's Blastx program (http://www.ncbi.nlm.nih.gov/blast) against the viral RefSeq database. To further confirm the results, selected contigs related to viruses were individually analysed using Blastx against the GenBank database.

Confirmation of infections The presence of all viruses identified by HTS was confirmed in the same samples by other diagnostic methods. Total RNA extractions were performed using the adsorption of nucleic acids on silica particles from 1 g of petioles or veins of leaves or scrapings of mature stems (Rott and Jelkmann 2001), grinding plant tissues in liquid

Table 1 Samples and overall results of sequence recovery representing viruses detected from grapevine (*Vitis* spp.) samples via high-throughput sequencing

Sample	Place of collection[a]	Host species	Cultivar	Virus-like symptoms	Yield (Mbp)	Number of assembled contigs	Contigs with virus hits	Viruses identified[c]
1 M	RS	*Vitis vinifera*	Cabernet Sauvignon	strong symptomatic	2057 (20.356.638)[b]	1374	184	GCSV, GVCV, GRGV, GLRaV-2, GLRaV-3, GRSPaV, GVA, GVB
2 M	RS	*V. flexuosa*	wild grapevine	asymptomatic	1683 (16.654.870)[b]	1589	51	GRGV, GSyV-1, GLRaV-2, GLRaV-3, GRSPaV, GFkV, GRVFV
1	RS	*V. labrusca*	Isabel	mild symptomatic	13341	2417	74	GLRaV-2, GLRaV-3, GRSPaV, GVA, GVB
3	SP	*V. labrusca*	Isabel	mild symptomatic	8070	2425	4	GRSPaV
4	SP	*V. gigas*	wild grapevine	asymptomatic	7425	3054	19	GRSPaV
5	SP	*V. vinifera*	Red Meire	strong symptomatic	5342	2424	110	GRGV, GLRaV-3, GRSPaV, GVA, GVB, GVD, GFkV
6	SP	*V. vinifera*	Moscato de Hamburgo	strong symptomatic	5829	2544	42	GRGV, GSyV-1, GLRaV-4, GRSPaV, GRVFV
7	PE	*V. vinifera*	Syrah	strong symptomatic	7773	3743	17	GRSPaV
8	PE	*V. vinifera*	Tempranillo	strong symptomatic	7550	2685	2	GRSPaV
9	RS	*V. tillifolia*	wild grapevine	asymptomatic	509	876	4	GRSPaV
10	RS	*V. vinifera*	Italia	strong symptomatic	1285	690	81	GRGV, GLRaV-2, GLRaV-3, GRSPaV, GVB, GFkV
12	RS	*V. vinifera*	CG 90450	strong symptomatic	6584	1553	33	GLRaV-2, GRSPaV
17	RS	*V. labrusca*	Tardia de Caxias	mild symptomatic	9545	3017	90	GLRaV-3, GRSPaV, GFkV
18	RS	*V. vinifera*	Trajadura	strong symptomatic	9759	2583	99	GRGV, GSyV-1, GLRaV-3, GLRaV-4 strain 5, GRSPaV, GVA, GFkV, GRVFV
19	RS	*V. vinifera*	Cabernet Franc	strong symptomatic	9343	1915	52	GRGV, GLRaV-2, GRSPaV, GFkV, GRVFV
Total					96095	32889	862	

[a] Brazilian states of Rio Grande do Sul (RS), São Paulo (SP) and Pernambuco (PE)

[b] Sequence reads

[c] GCSV, Grapevine Cabernet Sauvignon reovirus; GFkV, Grapevine fleck virus; GLRaV-2, Grapevine leafroll-associated virus 2; GLRaV-3, Grapevine leafroll-associated virus 3; GLRaV-4, Grapevine leafroll-associated virus 4; GRGV, Grapevine Red Globe virus; GRSPaV, Grapevine rupestris stem pitting-associated virus; GRVFV, Grapevine rupestris vein feathering virus; GSyV-1, Grapevine Syrah virus 1; GVA, Grapevine virus A; GVB, Grapevine virus B; GVCV, Grapevine vein clearing virus; GVD, Grapevine virus D

nitrogen. Total RNA or dsRNA extracted from fresh plant material from the original source, including healthy and positive controls, were analyzed by real-time RT-PCR (RT-qPCR) or conventional RT-PCR, aiming to confirm identifications of viral species infecting the evaluated grapevines. To confirm infection by DNA viruses, total DNA was extracted from infected samples using the DNeasy Plant Mini kit (Qiagen). Thus, all samples were indexed for the HTS identified viruses.

RT-qPCR reactions (One Step RT-qPCR) were carried out in 96-well plates using the TaqMan Fast Virus 1-Step Master Mix kit (Life Technologies) as described by Fajardo et al. (2016). The reaction data were analyzed as presence/absence assays and graphically, using the StepOne Software v.2.3 (Applied Biosystems), by determining the quantification cycle (Cq). Specific primers and Tamra probes used for virus detection by RT-qPCR have been previously described (Osman et al. 2007; Osman and Rowhani 2008; Bianchi et al. 2015;

Catarino et al. 2015). The presence of viruses recently detected or hitherto undetected in Brazil was confirmed by conventional RT-PCR, using the following primers: Ctg468F/Ctg468R for grapevine Cabernet Sauvignon reovirus (GCSV) (Al Rwahnih et al. 2015), RG-CF-F1/RG-CF-R1 for *Grapevine Red Globe virus* (GRGV) (Beuve et al. 2015), GSyV-1Det-F/GSyV-1Det-R for *Grapevine Syrah virus 1* (GSyV-1) (Al Rwahnih et al. 2009) and GVCV4142F/GVCV4387R for *Grapevine vein clearing virus* (GVCV) (Guo et al. 2014).

RT-PCR in a single step was carried out using the One Step RT-PCR kit (Qiagen) and reactions were performed according to the manufacturer with 4 μL (ca. 400 ng) of total RNA. Also aiming to confirm the virus identification by HTS, some obtained amplicons were purified, ligated into the pGEM-T-Easy vector (Promega) and the recombinant plasmids from transformed bacterial colonies were sequenced by the Sanger method (Fajardo et al. 2016).

Nucleotide and deduced amino acid sequences obtained by HTS for all virus species were deposited in GenBank and aligned with the reference sequences of these viruses as well as with the more similar sequences of related viruses available in GenBank using the NCBI's Blastn program (http://www.ncbi.nlm.nih.gov/blast). The GenBank accession codes of the nucleotide sequences of the isolates used for sequence analyses are presented in Tables 2 and 3.

Grafting assays One original source, cultivar Cabernet Sauvignon (Table 1), selected among other infected samples, was chosen for a graft-transmissibility assay aimed at observing probable specific symptoms in the cv. 1103P (*Vitis berlandieri* x *V. rupestris*), the main rootstock used in the temperate Brazilian grape-growing area. Sixteen plants of cv. 1103P were grafted with buds of multiple virus-infected Cabernet Sauvignon. After 7 months, symptoms were evaluated and plants were assayed by PCR using specific primer pairs to GVCV (Guo et al. 2014) as described above.

Description of the viral genetic variability and selection analysis The sequences generated in this study and additional sequences available in GenBank were used in the analysis of genetic variability (Supplementary Table S1). Only data sets of viral species with at least four sequences were used. The molecular variability descriptors [total number of segregating sites (S), mean nucleotide differences between sequences (K), nucleotide diversity (π), haplotype number (H), haplotype diversity (Hd) and Watterson's estimator for the population-scaled mutation rate] were estimated using DnaSP software v.5.10 (Rozas et al. 2003). The mean pairwise number of nucleotide diversity per site was also calculated using a sliding window of 100 bases, with a step size of 25 bases across different coding regions of the following viral species: *Grapevine fleck virus* (GFkV), *Grapevine leafroll-associated virus 2* (GLRaV-2), *Grapevine leafroll-associated virus 3* (GLRaV-3), *Grapevine leafroll-associated virus 4* (GLRaV-4), GLRaV-4 strain 5, *Grapevine Red Globe virus* (GRGV), *Grapevine rupestris stem pitting-associated virus* (GRSPaV), *Grapevine Syrah virus 1* (GSyV-1), *Grapevine virus A* (GVA) and *Grapevine virus B* (GVB). Mean values of non-synonymous to synonymous substitution ratios (dN/dS) for the different coding regions from grapevine viruses were analyzed using the Single Likelihood Ancestor Counting (SLAC) method within the HyPhy software (http://www.hyphy.org) implemented in the Datamonkey

Table 2 Virus sequences obtained by HTS from grapevine (*Vitis* spp.) in Brazil and used to assess genetic variability

Virus	Length (nt)	Coverage	GenBank access number	Grapevine species/cultivar and isolate name
GCSV	1110 up to 3849	near-complete genome (10 segments)	KR107527-KR107536	Cabernet Sauvignon (CS-BR)
GFkV	693	coat protein	KX828706	Italia (IT-BR)
GLRaV-2	16491	complete genome	KX774192	Isabel n.1 (ISA-BR)
GLRaV-3	18019	complete genome	KX701860	Isabel n.1 (ISA-BR)
GLRaV-3	18498	complete genome	KX756668	Tardia de Caxias (TC-BR)
GLRaV-3	18313	complete genome	KX756669	Trajadura (TRAJ-BR)
GLRaV-4	819	coat protein	KX828707	Moscato de Hamburgo (MH-BR)
GLRaV-4 strain 5	13823	complete genome	KX828702	Trajadura (TRAJ1-BR)
GRGV	6754	partial genome	KX828704	Cabernet Franc (CF-BR)
GRSPaV	8716	partial genome	KT948710	*V. flexuosa* (VF1)
GRSPaV	8659	complete ORFs	KX925555	Isabel (ISA-BR)
GRSPaV	8700	complete ORFs	KX925556	Tempranillo (TEMP-BR)
GRSPaV	8743	complete genome	KX958435	Cabernet Sauvignon (CS-BR)
GRVFV	6549	polyprotein gene	KX828705	Trajadura (TRAJ3-BR)
GSyV-1	6438	near-complete genome	KR153306	*V. flexuosa* (VF-BR)
GSyV-1	6440	partial polyprotein	KT037017	Moscato de Hamburgo (MH)
GSyV-1	6465	near-complete genome	KX130754	Trajadura (TRAJ-BR)
GVA	7341	partial genome	KX828703	Trajadura (TRAJ2-BR)
GVB	7605	complete genome	KX790785	Isabel n.1 (ISA-BR)
GVD	780	coat protein and RNA binding protein	KX828708	Red Meire (RM-BR)
GVCV	472	partial polyprotein	KR107537	Cabernet Sauvignon (CS-BR)

Table 3 Percent identities among Brazilian isolates of grapevine viruses obtained by HTS and homologous sequences retrieved from Genbank used for pairwise comparisons

Virus	Isolate name (GenBank access number)	Isolate used for comparison (country of origin, GenBank access number)	% nucleotide identity	% deduced amino acid identity
GCSV	CS-BR (KR107529)	LV89-15 (USA, KM378720)	97	99 (RNA polymerase)
	CS-BR (KR107532)	LV89-15 (USA, KM378725)	96	98 (CP)
GFkV	IT-BR (KX828706)	MT48 (Italy, AJ309022)	97	100 (CP)
		BF (Brazil, JN022610)	96	99 (CP)
		Cl1155 (Chile, HQ688989)	94	100 (CP)
GLRaV-2	ISA-BR (KX774192)	SG (USA, KF220376)	90	90 (RdRP) 95 (CP)
		93/955 (South Africa, AY881628)	86	88 (RdRP) 94 (CP)
		3138-07 (Canada, JX559644)	86	87 (RdRP) 95 (CP)
		OR2 (USA, JQ771955)	86	87 (RdRP) 95 (CP)
		OR1 (USA, FJ436234)	86	87 (RdRP) 95 (CP)
GLRaV-3	TRAJ-BR (KX756669)	3138-07 (Canada, JX559645)	99	99 (RdRP) 100 (CP)
		WA-MR (USA, GU983863)	99	99 (RdRP) 100 (CP)
		621 (South Africa, GQ352631)	99	99 (RdRP) 100 (CP)
		Cl-766 (Chile, EU344893)	99	98 (RdRP) 100 (CP)
		NY1 (USA, AF037268)	98	97 (RdRP) 99 (CP)
		623 (South Africa, GQ352632)	93	95 (RdRP) 96 (CP)
		GP18 (South Africa, EU259806)	92	94 (RdRP) 96 (CP)
GLRaV-4	MH-BR (KX828707)	LR106 (USA, FJ467503)	99	99 (CP)
		SS-N (India, KJ542647)	99	98 (CP)
		Y252-IL (Israel, AM176759)	92	94 (CP)
		Y253-TK (Turkey, AM162279)	90	93 (CP)
GLRaV-4 strain 5	TRAJ1-BR (KX828702)	3138-03 variant (Canada, JX559640)	95	96 (RdRP) 95 (CP)
		GLAV5 1050–02 (Canada, JX513893)	93	98 (RdRP) 95 (CP)
		3138-03 (Canada, JX559639)	93	96 (RdRP) 95 (CP)
		Y217 (France, FR822696)	93	96 (RdRP) 95 (CP)
GRGV	CF-BR (KX828704)	Graciano-T101 (Spain, KX171166)	88	95 (replicase polyprotein)
		Graciano-T90 (Spain, KX109927)	88	94 (replicase polyprotein)
		Graciano-T53 (Spain, KX171167)	88	94 (replicase polyprotein)
GRSPaV	VF1 (KT948710)	GRSPaV-JF (China, KR054734)	90	93 (ORF1) 98 (CP)
		BS (Canada, AY881627)	85	85 (ORF1) 98 (CP)
		RSPaV-PN (USA, AY368172)	84	90 (ORF1) 97 (CP)
		Syrah (USA, AY368590)	80	85 (ORF1) 93 (CP)
GRVFV	TRAJ3-BR (KX828705)	- (Italy, AY128949)	82	93 (polyprotein)
		- (USA, AY706994)	77	87 (polyprotein)
GSyV-1	TRAJ-BR (KX130754)	- (USA, FJ436028)	97	99 (polyprotein) 92 (MP)
		MG-02 (USA, FJ977041)	97	98 (polyprotein) 91 (MP)
		GSV1 3138–01 (USA, JX513896)	96	98 (polyprotein) 92 (MP)
		SK30 (Slovakia, KP221256)	93	97 (polyprotein) 78 (MP)
		CZ10 (Czech Republic, KP221255)	93	97 (polyprotein) 91 (MP)
GVA	TRAJ2-BR (KX828703)	I327-5 (South Africa, KC962564)	89	93 (MP) 98 (CP)
		GTG11-1 (South Africa, DQ855084)	82	95 (MP) 98 (CP)
		PA3 (Israel, AF007415)	82	84 (MP) 99 (CP)
		- (Israel, AY244516)	82	84 (MP) 96 (CP)
GVB	ISA-BR (KX790785)	GVB-QMWH (China, KF700375)	80	91 (ORF1) 96 (CP)
		94/971 (South Africa, EF583906)	77	87 (ORF1) 95 (CP)
		GVB-H1 (South Africa, GU733707)	77	86 (ORF1) 95 (CP)

Table 3 (continued)

Virus	Isolate name (GenBank access number)	Isolate used for comparison (country of origin, GenBank access number)	% nucleotide identity	% deduced amino acid identity
		GVB 953–1 (South Africa, KJ524452)	77	86 (ORF1) 93 (CP)
GVD	RM-BR (KX828708)	- (Italy, Y07764)	84	92 (CP)
		Garg (Brazil, JQ031716)	84	90 (CP)
		Dolc (Brazil, JQ031715)	84	90 (CP)
GVCV	CS-BR (KR107537)	LBC0903 (USA, JF301669)	85	97 (partial polyprotein)
		VRU2 (USA, KT907478)	85	97 (partial polyprotein)
		VRU (USA, KJ725346)	83	97 (partial polyprotein)

server (www.datamonkey.org), with default parameters. The nucleotide substitution models used are described in Supplementary Table S2. To avoid the effect of recombination events on selection analysis, recombination analysis using the program RDP v.4 (Martin et al. 2015) and the GARD method (available at the Datamonkey server) was performed. In recombination analysis using the program RDP, only recombination events detected by at least three of the methods available in the program were considered reliable. Recombinants detected in the RDP program were excluded from the selection analysis.

Results and discussion

HTS Plant viral metagenomic approaches have targeted five main classes of nucleic acids to maximize the proportion of virus-derived sequence reads obtained from HTS runs: (*i*) dsRNA, (*ii*) total RNA or (*iii*) DNA, (*iv*) siRNA or (*v*) virion-associated nucleic acids purified from virus particles. RNA viruses generate dsRNA during the process of replication. Double-stranded RNA has provided an in-depth analysis of virus-specific sequences. Moreover, it allows enrichment of virus-specific RNA in the samples, since virus particle purification from grapevines is often problematic and cumbersome (Roossinck et al. 2015).

HTS analysis resulted in detection of multiple virus species in mixed infections in analysed grapevines (Table 1). We were able to identify assembled contigs of 14 viruses, some of them covering large extensions of the viral genomes including some near-complete and complete genomes which were obtained for GCSV, GLRaV-2, −3 and −4 strain 5, GRSPaV, GSyV-1 and GVB (Table 2). At least one sequence per virus species detected by HTS was deposited in GenBank (Table 2).

In general, the following viruses were found in mixed infections: GCSV, family *Reoviridae*; GFkV and GRGV (*Tymoviridae*, genus *Maculavirus*); GLRaV-2 (*Closteroviridae*, *Closterovirus*); GLRaV-3, GLRaV-4 and GLRaV-4 strain 5 (*Closteroviridae*, *Ampelovirus*); GRSPaV

(*Betaflexiviridae*, *Foveavirus*); *Grapevine rupestris vein feathering virus* (GRVFV) and GSyV-1 (*Tymoviridae*, *Marafivirus*); GVA, GVB and *Grapevine virus D* (GVD) (*Betaflexiviridae*, *Vitivirus*); and GVCV (*Caulimoviridae*, *Badnavirus*) (Table 1). Ten viruses had been already reported, frequently infecting grapevines in Brazil (GLRaV-2, −3, −4, −4 strain 5, GVA, GVB, GVD, GRSPaV, GFkV, GRVFV) (Basso et al. 2017; Catarino et al. 2015), three had not been previously reported in this country (GCSV, GRGV and GVCV) and one had only recently been identified (GSyV-1) (Moura et al. 2017). It is interesting to highlight the association of some viruses with wild grapevine genotypes (Table 1). Further analysis will be necessary to better establish the prevalence of these detected viruses in productive Brazilian vineyards and to evaluate their potential effects on grape yield, juice and wine quality, as well as to study their epidemiology, including possible interactions among grapevine viruses.

GCSV was discovered by HTS as a novel virus species infecting grapevines in USA (Al Rwahnih et al. 2015) and, to our knowledge, Brazil is only the second country where GCSV has been reported. GSyV-1 was first identified in USA (Al Rwahnih et al. 2009) and symptoms include swollen graft unions, cracking and pitting of the wood, stem necrosis, leaf reddening and scorching, vine decline and death. Its presence has been reported in many countries (Glasa et al. 2015) and was recently reported infecting vineyards in Brazil (Moura et al. 2017). GVCV was found associated with leaf vein-clearing, shortening of the internodes, crinkled, misshapen and smaller leaves and reduced grape berry size, with irregular shape and abnormal texture (Guo et al. 2014; Zhang et al. 2011) and GRGV was previously reported in many viticultural countries (Beuve et al. 2015).

The vegetative propagation of grapevine and vector transmission favor the spread and promotes the accumulation of different virus species in the same plant, which, depending on the genotype, may exhibit symptoms or remain symptomless (Maliogka et al. 2015). Thus, the capacity to detect asymptomatic viral infections and coinfections showed by HTS technologies is emphasized.

In the present study, HTS made possible to draw a comprehensive picture of viruses infecting grapevines in Brazil (Table 1). Double-stranded RNA as template for library preparation was a suitable option for the identification of the viruses present in the plants; comparative analysis disclosed the prevailing presence of GRSPaV (15 infected samples), GLRaV-3 and GRGV (7), GFkV and GLRaV-2 (6) (Table 1). This fact was presumably due to the propensity of these viruses to accumulate RNA replicative intermediates (dsRNA) in mature grapevine canes collected in winter (Martin et al. 2011).

As for GRSPaV, the most prevalent virus found in the analysed samples, some of its characteristics and properties help to explain this finding. Dissemination and transmission of GRSPaV are through the use of infected propagating materials, also there is the possibility of transmission through pollen and seeds and no natural vector is known. Natural infection of GRSPaV is restricted to grapevines (*Vitis* spp.) and it has been detected in many cultivated *V. vinifera* varieties, as well as in *V. sylvestris*, *V. rupestris*, *V. riparia* and French-American hybrids. GRSPaV is an ancient virus and has co-existed with grapevines since antiquity. The ancestor of present day GRSPaV gained entry into different *Vitis* species including wild species at some point in the remote past, and as a result of co-evolution and adaptation to different *Vitis* species, the genome of the ancestral virus diverged significantly (Meng and Gonsalves 2007).

Because dsRNA is synthesized by RNA viruses and viroids as replicative intermediates and plants do not normally produce dsRNA, sequencing total dsRNA dramatically increases the proportion of reads specific to viruses and viroids (Wu et al. 2015). A study compared deep sequencing of total RNA and dsRNA from the same plant samples and found that virus reads increased from 2 to 53% after dsRNA enrichment (Al Rwahnih et al. 2009). However, dsRNA-based analyses present some shortcomings such as their reduced effectiveness for the detection of negative sense ssRNA viruses, since these do not accumulate large amounts of dsRNA during replication (Roossinck et al. 2015). Also, Wu et al. (2015) mentioned that six of the seven new viruses discovered by dsRNA sequencing contain RNA genomes, possibly because plant DNA viruses do not produce sufficiently long dsRNA in their life cycle and just one new DNA geminivirus was identified from sequencing a total dsRNA preparation (Al Rwahnih et al. 2013). In this work it was possible to obtain four short contigs (274–482 bp) of a dsDNA virus, a *Badnavirus* from a Cabernet Sauvignon sample (Tables 1 and 2).

Confirmation of infections All grapevine virus infections (by GCSV, GFkV, GLRaV-2, −3, −4 and −4 strain 5, GRGV, GRSPaV, GRVFV, GSyV-1, GVA, GVB, GVD and GVCV) as determined by HTS (Table 1) were confirmed by RT-qPCR (data not shown) and/or (RT-)PCR/Sanger sequencing from the same analysed samples using specific reagents in the

performed reactions. Recombinant plasmids containing cloned amplicons of GCSV (386 bp, GenBank KR074408), GRGV (453 bp, KY039178), GSyV-1 (296 bp, KU258395) and GVCV (246 bp, KY039180) were sequenced. Nucleotide identities were very high when compared with corresponding sequences generated by HTS and with other related foreign isolates [98% with KM378723 (GCSV), 95% with KM491303 (GRGV), 97% with JN698963 (GSyV-1) and 91% with JF301669 (GVCV)], thus confirming the virus detections by HTS in the analysed samples. Nucleotide and deduced amino acid identities among obtained viral contigs and related grapevine virus sequences in GenBank were high (77% to 99%), being higher when conserved ORFs/genes such as the coat protein were considered for comparison (Table 3).

Grafting assays Graft-transmissibility of multiple virus-infected buds of cv. Cab. Sauvignon onto rootstock '1103P' resulted in the expression of symptoms in 12 out of 16 grafted grapevines. Seven months after grafting, '1103P' leaves exhibited vein-clearing, leaf distortion and few necrotic spots along the major veins (Fig. 1). These symptoms are very similar to those reported by Beach et al. (2017). These authors mentioned that symptoms of GVCV-infected wild *V. rupestris* grapevine appeared initially as translucent vein clearing on young leaves and progressed to vein necrosis on mature leaves. Healthy control grapevines showed no symptoms. It is important to highlight that the original source of the buds (cv. Cab. Sauvignon) was infected with at least eight viruses as determined by HTS (Table 1, Fig. 1). This viral complex might have influenced symptom expression, but nevertheless it was still possible to observe symptoms which were similar to those previously reported for GVCV. All grafted '1103P' plants exhibiting symptoms were tested for GVCV by PCR using specific primer pairs and the presence of the virus was confirmed. The economic importance of GVCV has not yet been clearly defined (Zhang et al. 2011).

Description of viral genetic variability and selection analysis The molecular variability of GFkV, GLRaV-2, GLRaV-3, GLRaV-4, GLRaV-4 strain 5, GRGV, GRSPaV, GSyV-1, GVA and GVB was evaluated using different coding regions. The capsid protein genes always showed lower genetic variability than other analized coding regions (Supplementary Table S3). Descriptors for MP of GVA and ORF1 (replicase) of GRSPaV and GVB isolates indicated higher genetic variability when compared with other coding regions of the analyzed viruses, represented by a higher number of segregating sites (S), nucleotide diversity (π), haplotype number (H) and haplotype diversity (Hd) (Supplementary Table S3).

Nucleotide diversity (π) was lower than 0.2288 for all viruses in different coding regions (Supplementary Table S3). The virus GVB (ORF1 - replicase) had the higher π value (0.22880

Fig. 1 **a** Healthy, asymptomatic grapevine cv. Cabernet Sauvignon. **b** Cultivar Cabernet Sauvignon, original source, infected with eight viruses as determined by HTS and (RT-)PCR, showing severe symptoms. **c** Healthy *Vitis berlandieri* x *V. rupestris* '1103P'. **d** *Vitis berlandieri* x *V. rupestris* '1103P' grafted with cv. Cabernet Sauvignon showing similar symptoms of *Grapevine vein clearing virus* (GVCV) including vein-clearing, leaf distortion and necrotic spots along the major veins, 7 months post-grafting
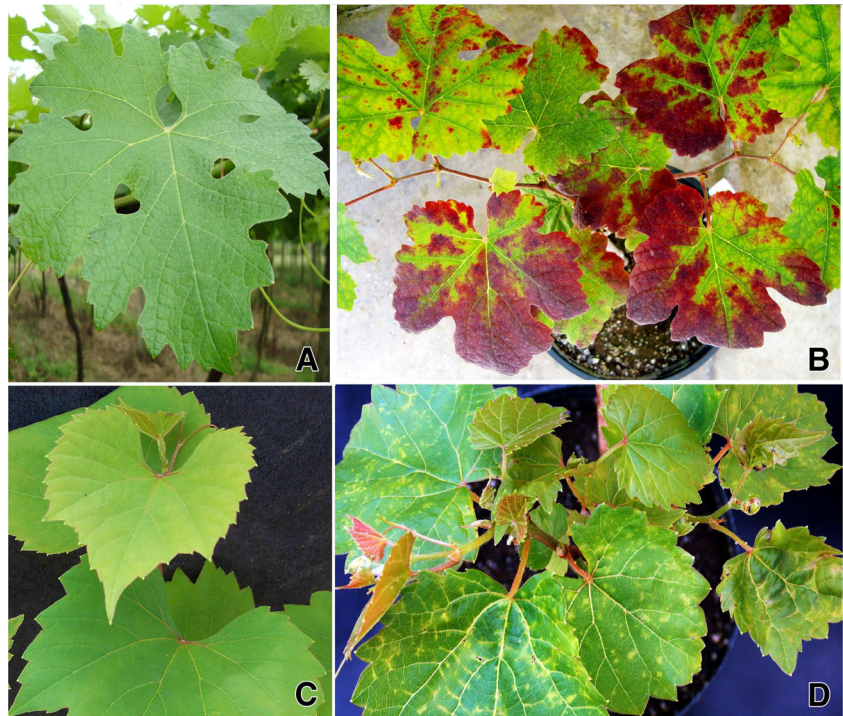
±0.02406). The lowest π value was detected in GLRaV-3 coat protein gene (0.03846 ± 0.00780), consistent with previous work (Moura et al. 2017). In general, the π values detected in this work are in accordance with π values found by other authors in studies involving viruses with DNA and RNA genomes (Garcia-Arenal et al. 2001; Lima et al. 2013; Zanardo et al. 2014a; Moura et al. 2017). Nucleotide diversity (π) was also analyzed along the coding regions of GFkV, GLRaV-2, GLRaV-3, GLRaV-4, GLRaV-4 strain 5, GRGV, GRSPaV, GSyV-1, GVA and GVB. The behavior for π values of the coat protein gene in all viruses analyzed was similar to each other, differing only in the absolute values of π (Fig. 2). The peaks for nucleotide diversity values in the coding regions of the different viruses were found predominantly in the N-terminal and central region and were not evenly distributed along the genome (Fig. 2). Watterson's estimator for the population-scaled mutation rate (Θ-W) for most of the analyzed viruses was in the order of 10^{-2} (except for the RdRP of GLRaV-2 and GLRaV-4 strain 5; CP of GLRaV-4; partial polyprotein of GRGV; ORF1 of GRSPaV and GVB; and MP of GVA and ORF 1, for which it was in the order of 10^{-1}) (Supplementary Table S3). The Θ-W values of 10^{-1} are lower when compared with other coding regions and other viral species described in the literature (Lima et al. 2013; Moura et al. 2017; Rocha et al. 2013; Zanardo et al. 2014a). It is noteworthy that datasets with few sequences may mask the results of variability descriptors, over- or underestimating the genetic variability.

Recombination events interfere with the selection analysis. Thus, recombination was analyzed and thirty-seven recombination events were detected (Supplementary Table S4). Except

for GLRaV-2, GLRaV-4 strain 5 and GRGV, all viruses analyzed in this work have at least one recombination event. Of the total number of recombination events, six are related to sequences characterized by HTS in this study [recombinants IT-BR, KX828706 (CP of GFkV); ISA-BR, KX925555 and TEMP-BR, KX925556 (both CP of GRSPaV); CS-BR, KX958435 (ORF1 of GRSPaV); and VF-BR, KR153306 (Polyprotein of GSyV-1)] (Supplementary Table S4). Although the number of GSyV-1 sequences used in the analysis is small, this virus showed the greatest number of recombination events (Supplementary Table S4). The results of the recombination analysis suggest that this variability mechanism was important for generating genetic variation of the grapevine viruses analyzed in this work, in agreement with several studies that have shown the importance of recombination in the evolution of plant viruses (Fan et al. 2015; Farooq et al. 2013; Garcia-Arenal et al. 2001; 2003; Lima et al. 2013; Simon-Loriere and Holmes 2011; Rocha et al. 2013; Zanardo et al. 2014b). All coding regions of the analyzed viruses showed dN/dS ratios (non-synonymous/synonymous substitutions ratios) lower than 1.0, indicating negative or purifying selection (Table 4). The coat protein gene of GFkV showed the lowest dN/dS ratio (0.0189). Selection pressures can be associated with the maintenance of structural features of the virus. Viral proteins are multifunctional and may be involved in many steps

Fig. 2 Mean pairwise number of nucleotide differences per site ▶ (nucleotide diversity, π) calculated on a sliding window across different coding regions of grapevine viruses
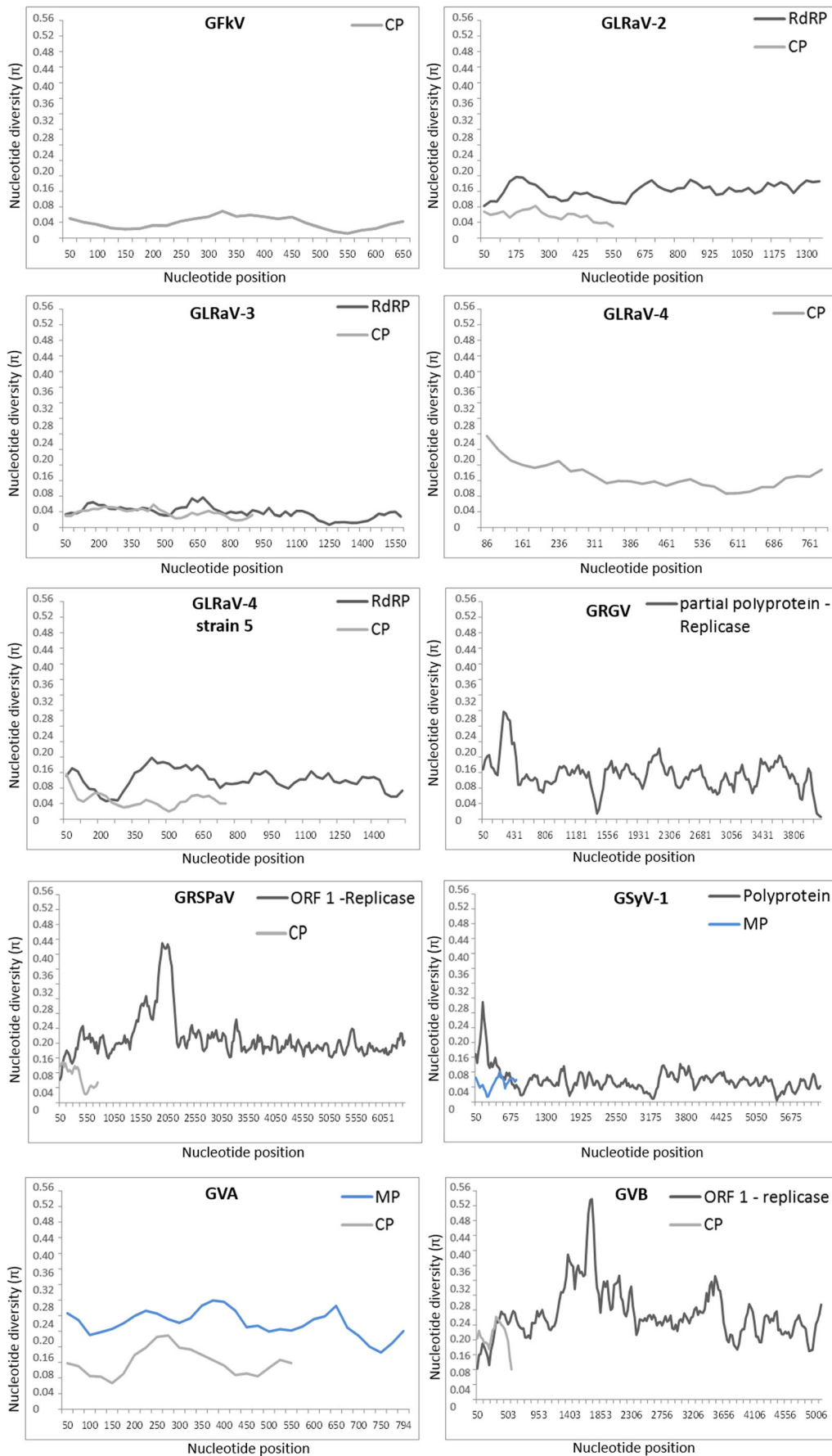
Table 4 Mean values of non-synonymous to synonymous substitution ratios (d_N/d_S) for different coding regions of grapevine viruses

Virus[a]	Genome region	Number of sequences	d_N/d_S
GFkV	CP	6	0.0189
GLRaV-2	CP	95	0.1738
	RdRP	12	0.0436
GLRaV-3	CP	98	0.2699
	RdRP	13	0.1046
GLRaV-4	CP	6	0.1159
GLRaV-4 strain 5	CP	76	0.2741
	RdRP	5	0.0652
GRGV	partial polyprotein	4	0.0418
GRSPaV	CP	88	0.1029
	ORF1	16	0.0952
GSyV-1	MP	8	0.0922
	polyprotein	8	0.0701[b]
GVA	CP	82	0.0453
	MP	14	0.1105
GVB	CP	61	0.0920
	ORF1	5	0.0724

CP coat protein; *MP* movement protein; *ORF1* open reading frame 1 (replicase); *RdRP* RNA-dependent RNA polymerase

[a] GFkV, *Grapevine fleck virus*; GLRaV-2, *Grapevine leafroll-associated virus 2*; GLRaV-3, *Grapevine leafroll-associated virus 3*; GLRaV-4, *Grapevine leafroll-associated virus 4*; GRGV, *Grapevine Red Globe virus*; GRSPaV, *Grapevine rupestris stem pitting-associated virus*; GSyV-1, *Grapevine Syrah virus 1*; GVA, *Grapevine virus A*; GVB, *Grapevine virus B*

[b] Recombinant sequences were not excluded, because the restricted number of sequences would interfere with the analysis

of viral infection processes. Thus the negative selection, predominant in the coding regions of viral proteins (Garcia-Arenal et al. 2001), acts to prevent damaging changes to vital functional domains.

The identification and discovery of grape viruses by HTS strengthens the argument for the incorporation of this technology into grapevine indexing programs for the detection of viral pathogens of agronomic significance. This is especially the case of unclear symptoms and viruses which presence would be undetectable by biological indexing or that could escape detection by RT-qPCR assays since these require molecular knowledge of the viral pathogen (Al Rwahnih et al. 2015). In conclusion, the range of information generated in this study confirms the huge potential of HTS technology to expand the knowledge of viral infections in grapevines.

References

Armijo G, Schlechter R, Agurto M, Muñoz D, Nuñez C, Arce-Johnson P (2016) Grapevine pathogenic microorganisms: understanding infection strategies and host response scenarios. Front Plant Sci 7:382

Al Rwahnih M, Daubert S, Golino D, Rowhani A (2009) Deep sequencing analysis of RNAs from a grapevine showing Syrah decline symptoms reveals a multiple virus infection that includes a novel virus. Virology 387:395–401

Al Rwahnih M, Dave A, Anderson MM, Rowhani A, Uyemoto JK, Sudarshana MR (2013) Association of a DNA virus with grapevines affected by red blotch disease in California. Phytopathology 103:1069–1076

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

Basso MF, Fajardo TVM, Saldarelli P (2017) Grapevine virus diseases: economic impact and current advances in viral prospection and management. Rev Bras de Frutic 39, in press

Beach S, Kovens M, Hubbert L, Honesty S, Guo Q, Pap D, Dai R, Kovacs L, Qiu W (2017) Genetic and phenotypic characterization of *Grapevine vein clearing virus* from wild *Vitis rupestris*. Phytopathology 107:138–144

Beuve M, Candresse T, Tannières M, Lemaire O (2015) First report of *Grapevine redglobe virus* (GRGV) in grapevine in France. Plant Dis 99:422

Bianchi GL, de Amicis F, de Sabbata L, di Bernardo N, Governatori G, Nonino F, Prete G, Marrazzo T, Versolatto S, Frausin C (2015) Occurrence of *Grapevine Pinot gris virus* in Friuli Venezia Giulia (Italy): field monitoring and virus quantification by real-time RT-PCR. EPPO Bull 45:22–32

Burger JT, Maree HJ (2015) Metagenomic next-generation sequencing of viruses infecting grapevines. Methods Mol Biol 1302:315–330

Catarino AM, Fajardo TVM, Pio-Ribeiro G, Eiras M, Nickel O (2015) Incidência de vírus em videiras no Nordeste brasileiro e caracterização molecular parcial de isolados virais locais. Ciência Rural 45:379–385

Coetzee B, Freeborough M-J, Maree HJ, Celton J-M, Rees DJG, Burger JT (2010) Deep sequencing analysis of viruses infecting grapevine: virome of a vineyard. Virology 400:157–163

Fajardo TVM, Eiras M, Nickel O (2016) Detection and molecular characterization of *Grapevine yellow speckle viroid 1* isolates infecting grapevines in Brazil. Trop Plant Pathol 41:246–253

Fan X, Hong N, Dong Y, Ma Y, Zhang ZP, Ren F, Hu G, Zhou J, Wang G (2015) Genetic diversity and recombination analysis of grapevine leafroll-associated virus 1 from China. Arch Virol 160:1669–1678

Farooq AB, Ma YX, Wang Z, Zhuo N, Wenxing X, Wang GP, Hong N (2013) Genetic diversity analyses reveal novel recombination events in *Grapevine leafroll-associated virus 3* in China. Virus Res 171:15–21

Garcia-Arenal F, Fraile A, Malpica JM (2001) Variability and genetic structure of plant virus populations. Annu Rev Phytopathol 39:157–186

Garcia-Arenal F, Fraile A, Malpica JM (2003) Variation and evolution of plant virus populations. Int Microbiol 6:225–232

Giampetruzzi A, Roumi V, Roberto R, Malossini U, Yoshikawa N, La Notte P, Terlizzi F, Credi R, Saldarelli P (2012) A new grapevine virus discovered by deep sequencing of virus- and viroid-derived small RNAs in cv. Pinot gris. Virus Res 163:262–268

Glasa M, Predajna L, Soltys K, Sabanadzovic S, Olmos A (2015) Detection and molecular characterisation of *Grapevine Syrah virus-1* isolates from Central Europe. Virus Genes 51:112–121

Guo Q, Honesty S, Xu ML, Zhang Y, Schoelz J, Qui W (2014) Genetic diversity and tissue and host specificity of *Grapevine vein clearing virus*. Phytopathology 104:539–547

Hadidi A, Flores R, Candresse T, Barba M (2016) Next-generation sequencing and genome editing in plant virology. Front Microbiol 7:1325

Jo Y, Choi H, Cho JK, Yoon J-Y, Choi S-K, Cho WK (2015) *In silico* approach to reveal viral populations in grapevine cultivar Tannat using transcriptome data. Sci Rep 5:15841

Lima AT, Sobrinho RR, Gonzalez-Aguilera J, Rocha CS, Silva SJ, Xavier CA, Silva FN, Duffy S, Zerbini FM (2013) Synonymous site variation due to recombination explains higher genetic variability in begomovirus populations infecting non-cultivated hosts. J Gen Virol 94:418–431

Maliogka VI, Martelli GP, Fuchs M, Katis NI (2015) Control of viruses infecting grapevine. In: Loebenstein G, Katis NI (eds) Control of plant virus diseases: vegetatively-propagated crops, vol 91, Advances in Virus Research. Academic Press, Burlington, pp 175–227

Martelli GP (2014) Directory of virus and virus-like diseases of the grapevine and their agents. J Plant Pathol 96(suppl):S1–S136

Martin DP, Murrell B, Golden M, Khoosal A, Muhire B (2015) RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evol 1:vev003

Massart S, Olmos A, Jijakli H, Candresse T (2014) Current impact and future directions of high throughput sequencing in plant virus diagnostics. Virus Res 188:90–96

Martin RR, Jelkmann W, Tzanetakis IE (2011) Double-stranded RNAs and their use for characterization of recalcitrant viruses. In: Hadidi A, Barba M, Candresse T, Jelkmann W (eds) Virus and virus-like diseases of pome and stone fruits. APS Press, St. Paul, pp 323–326

Meng B, Gonsalves D (2007) *Grapevine rupestris stem pitting-associated virus*: a decade of research and future perspectives. Plant Vir 1:52–62

Moura CJM, Fajardo TVM, Eiras M, Silva FN, Nickel O (2017) Molecular characterization of GSyV-1 and GLRaV-3 and prevalence of grapevine viruses in a grape-growing area. Sci Agric 74, in press

Naidu RA, Marre HJ, Burger JT (2015) Grapevine leafroll disease and associated viruses: a unique pathosystem. Annu Rev Phytopathol 53:613–634

Osman F, Leutenegger C, Golino D, Rowhani A (2007) Real-time RT-PCR (TaqMan) assays for the detection of *Grapevine leafroll associated viruses* 1–5 and 9. J Virol Meth 141:22–29

Osman F, Rowhani A (2008) Real-time RT-PCR (TaqMan) assays for the detection of viruses associated with Rugose wood complex of grapevine. J Virol Meth 154:69–75

Poojari S, Alabi OJ, Fofanov VY, Naidu RA (2013) A leafhopper-transmissible DNA virus with novel evolutionary lineage in the family *Geminiviridae* implicated in grapevine redleaf disease by next-generation sequencing. PLoS ONE 8, e64194

Rocha CS, Castillo-Urquiza GP, Lima ATM, Silva FN, Xavier CAD, Hora-Júnior BT, Beserra-Júnior JEA, Malta AWO, Martin DP, Varsani A, Alfenas-Zerbini P, Mizubuti ESG, Zerbini FM (2013) Brazilian begomovirus populations are highly recombinant, rapidly evolving, and segregated based on geographical location. J Virol 87: 5784–5799

Roossinck MJ, Martin DP, Roumagnac P (2015) Plant virus metagenomics: advances in virus discovery. Phytopathology 105:716–727

Rott ME, Jelkmann W (2001) Characterization and detection of several filamentous viruses of cherry: adaptation of an alternative cloning method (DOP-PCR) and modification of an RNA extraction protocol. Eur J Plant Pathol 107:411–420

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497

Simon-Loriere E, Holmes EC (2011) Why do RNA viruses recombine? Nat Rev Microbiol 9:617–626

Valverde RA, Nameth ST, Jordan RL (1990) Analysis of double-stranded RNA for plant virus diagnosis. Plant Dis 74:255–258

Zanardo LG, Silva FN, Lima ATM, Milanesi DF, Castilho-Urquiza GP, Almeida AMR, Zerbini FM, Carvalho CM (2014a) Molecular variability of *Cowpea mild mottle virus* infecting soybean in Brazil. Arch Virol 159:727–737

Zanardo LG, Silva FN, Bicalho AAC, Urquiza GPC, Lima ATM, Almeida AMR, Zerbini FM, Carvalho CM (2014b) Molecular and biological characterization of *Cowpea mild mottle virus* isolates infecting soybean in Brazil and evidence of recombination. Plant Pathol 63:456–465

Wu Q, Ding S-W, Zhang Y, Zhu S (2015) Identification of viruses and viroids by next-generation sequencing and homology-dependent and homology-independent algorithms. Annu Rev Phytopathol 53:425–444

Zhang Y, Singh K, Kaur R, Qiu W (2011) Association of a novel DNA virus with the Grapevine vein-clearing and vine decline syndrome. Phytopathology 101:1081–1090