




Diverse genomoviruses representing twenty-nine species identified associated with plants

Rafaela S. Fontenele^{1,2} · Philippe Roumagnac^{3,4} · Cécile Richet^{3,4} · Simona Kraberger¹ · Daisy Stainton⁵ · Maketalena Aleamotu'a⁶ · Denis Filloux^{3,4} · Pauline Bernardo^{3,4,7} · Gordon W. Harkins⁸ · James McCarthy⁹ · Lachlan S. Charles¹⁰ · Natalia S. Lamas¹¹ · Emanuel F. M. Abreu¹¹ · Rayane A. Abreu^{11,12} · Graciete B. Batista^{11,12} · Ana L. M. Lacerda¹¹ · Andrew Salywon¹³ · Martin F. Wojciechowski² · Lucas C. Majure¹⁴ · Darren P. Martin¹⁵ · Simone G. Ribeiro^{11,12} · Pierre Lefeuvre¹⁶ · Arvind Varsani^{1,2,17} 

Received: 19 May 2020 / Accepted: 9 August 2020 / Published online: 6 September 2020
© Springer-Verlag GmbH Austria, part of Springer Nature 2020

Abstract

Genomoviruses (family *Genomoviridae*) are circular single-stranded DNA viruses that have been mainly identified through metagenomics studies in a wide variety of samples from various environments. Here, we describe 98 genomes of genomoviruses found associated with members of 19 plant families from Australia, Brazil, France, South Africa and the USA. These 98 genomoviruses represent 29 species, 26 of which are new, in the genera *Gemykolovirus* (n = 37), *Gemyduguivirus* (n = 9), *Gemygorvirus* (n = 8), *Gemykroznavirus* (n = 6), *Gemycircularvirus* (n = 21) and *Gemykibivirus* (n = 17).

The family *Genomoviridae* was established in 2016 [15] following the identification and characterization of the

Handling Editor: Elvira Fiallo-Olivé.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00705-020-04801-5>) contains supplementary material, which is available to authorized users.

✉ Arvind Varsani
arvind.varsani@asu.edu

- 1 The Biodesign Center for Fundamental and Applied Microbiomics and Center for Evolution and Medicine, Arizona State University, Tempe, AZ 85287-5001, USA
- 2 School of Life sciences, Arizona State University, Tempe, AZ 85287-5001, USA
- 3 CIRAD, BGPI, 34398 Montpellier, France
- 4 BGPI, INRAE, CIRAD, Institut Agro, Univ Montpellier, 34398 Montpellier, France
- 5 Department of Entomology and Plant Pathology, Division of Agriculture, University of Arkansas System, Fayetteville, AR 72701, USA
- 6 School of Environmental and Life Sciences, The University of Newcastle, Callaghan, NSW 2308, Australia
- 7 Enza Zaden, Haling 1-E, 1602 DB Enkhuizen, The Netherlands
- 8 South African MRC Bioinformatics Unit, South African National Bioinformatics Institute, University of the Western Cape, Bellville, South Africa

single-stranded DNA (ssDNA) virus *Sclerotinia sclerotiorum* hypovirulence-associated DNA virus 1 (SsHADV-1) [24] and further isolation of several similar sequences from a wide variety of samples [22] including plants [3, 5, 6, 9, 14, 16, 18, 19]. SsHADV-1 was found infecting and causing

- 9 Manaaki Whenua, Landcare Research, Lincoln 7640, New Zealand
- 10 Department of Botany and Plant Sciences, University of California, Riverside, CA 92507, USA
- 11 Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF, Brazil
- 12 PPG Ciências Naturais e Biotecnologia, Universidade Federal de Campina Grande, Cuité, PB, Brazil
- 13 Desert Botanical Garden, Phoenix, AZ 85008, USA
- 14 Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, USA
- 15 Computational Biology Division, Department of Integrative Biomedical Sciences, Institute of Infectious Diseases and Molecular Medicine, University of Cape Town, Observatory 7925, South Africa
- 16 CIRAD, UMR PVBMT, St. Pierre, 97410 La Réunion, France
- 17 Structural Biology Research Unit, Department of Clinical Laboratory Sciences, University of Cape Town, Observatory, 7925 Cape Town, South Africa

hypovirulence in the plant fungal pathogen *Sclerotinia sclerotiorum* and currently is the only genomovirus that has been experimentally associated with a host; the other isolated genomoviruses have no further host confirmation. The family *Genomoviridae* has been divided into nine genera based on the phylogeny of their replication-associated protein (Rep) amino acid sequence. These genera are *Gemycircularvirus*, *Gemyduguivirus*, *Gemygorvirus*, *Gemykibivirus*, *Gemykolovirus*, *Gemykrogvirus*, *Gemykroznavirus*, *Gemytondvirus* and *Gemyvongvirus* [22].

Genomovirus genomes are circular, between ~2.0 and 2.3 kb in length, and have at least three large open reading frames (ORFs). These ORFs encode a Rep on the complementary-sense genome strand that is likely expressed from a spliced transcript, a putative RepA protein on the complementary strand, and a putative capsid protein (CP) on the virion strand. The Reps of genomoviruses are evolutionarily most closely related to those of plant-infecting viruses of the family *Geminiviridae* [22]. Unlike geminiviruses, however, genomoviruses have no recognizable movement protein genes, and their capsid proteins (CPs) have no easily detectable homology to those of any other classified viruses [25].

While the vast majority of known plant viruses have been discovered in cultivated plants, less is known about the diversity and pathogenicity of viruses infecting non-cultivated plants (i.e., wild plants, including weeds) [23]. Additional information on the diversity of viruses associated with non-cultivated plants will provide critical new information on the ecology and evolution of plant viruses beyond the few cultivars tested to date.

In this study, we report the identification of 98 genomovirus genomes associated with 88 plant samples representing 19 plant families collected between 2008 and 2018 in Australia, Brazil, France, South Africa and the USA (Table 1). In the Western Cape Region of South Africa, plants were sampled randomly between 2010 and 2012 regardless of infection symptoms within a 4.5 km × 4.5 km plot [2]. In Northeast Brazil, both viral symptomatic and asymptomatic plants were collected between 2008 and 2015. In Australia, France and the USA, plants were opportunistically sampled regardless of symptoms. The genera and species of sampled plants were identified by expert botanists and from the sequences obtained by high-throughput sequencing. The virion-associated nucleic acids (VANA) method was used to process the samples collected in South Africa as described previously [2, 10]. Total DNA from the Australian, Brazilian, French and US samples was extracted using the Dellaporta method [7] and enriched for small circular DNA molecules by rolling-circle amplification (RCA) using TempliPhi (GE Healthcare, USA). RCA products were pooled according to sampling location and the taxonomic plant family from which the samples were obtained, and these pooled samples were then sequenced at Macrogen Inc. (South Korea) using the

Illumina HiSeq 2500 platform with 2 × 150 bp paired-end sequencing. The paired-end reads were assembled *de novo* using ABySS 2.0 [12], and contigs > 750 nt were compared to the viral GenBank sequence database using BLASTx [1]. Abutting primers were designed for contigs displaying detectable homology to known genomoviruses, and these were used to recover the full genomovirus genomes using PCR employing Kapa Hotstart HiFi polymerase (Kapa Biosystems, USA) and RCA products from the different plant samples as templates with the thermal cycler conditions recommended by Kapa Biosystems and an annealing temperature of 60 °C. Amplicons of ~2 kb were resolved on a 0.7% agarose gel, excised, and eluted using a MEGAquick-spin™ Total Fragment DNA Purification Kit (iNtRON Biotechnology, South Korea). The resulting DNA fragments were then cloned in pJET 1.2 (Thermo Fisher Scientific, USA), and the corresponding recombinant plasmid was sequenced by the Sanger method at Macrogen Inc. (South Korea). Finally, Geneious v11.1.4 [13] was used to assemble the sequence and annotate the genomovirus genome sequences.

All 416 full-length genomovirus genome sequences available in GenBank on the 1st of February 2020 were analyzed together with the 98 genome sequences determined in this study. Genome-wide pairwise nucleotide sequence identities, and pairwise amino acid sequence identities of Rep and CP were determined using SDT v1.2 [20]. The 98 genomovirus genomes represent 29 species (26 of which are new) based on the currently accepted genomovirus pairwise nucleotide sequence identity species demarcation threshold of 78% [22] (Supplementary Table 1 and Supplementary Table 2). The genomoviruses recovered in this study have been tentatively named "plant-associated genomovirus (PaGmV) 1 - 29". These new PaGmV sequences range in size from ~2.1-2 to 4 kb and, as with other known genomoviruses, have a conserved nonanucleotide within their intergenic regions that is the likely origin of virion-strand replication (Table 1). Also, as with other known genomovirus sequences, the new sequences also encode a probable CP on their presumed virion strand and both a putatively spliced Rep and a RepA on their presumed complementary strands.

A maximum-likelihood phylogenetic tree was inferred from the aligned Rep sequences using PHYML [11] with the rtRev+G amino acid substitution model (determined to be the best-fitting model by ProtTest [4]). Branches with approximate likelihood ratio test (aLRT) support <0.8 were collapsed using TreeGraph 2 [21]. Geminivirus Rep sequences were used to root the phylogenetic tree. The resulting phylogenetic tree shows that the PaGmVs could be classified as belonging to six out of the nine recognized genomovirus genera. PaGmV 7-10 (n = 37) were tentatively assigned to the genus *Gemykolovirus*, PaGmV 4-5 (n = 9) to the genus *Gemyduguivirus*, PaGmV 1 and 26 (n = 8) to the genus *Gemygorvirus*, and PaGmV 23-24 (n = 6) to the

Table 1 Information about the 98 genomoviruses recovered in this study, including their accession numbers, For each genome, the country, year of collection, associated plant genus and family are given. Also presented are the nonnucleotide sequence at the presumed origin of virion-strand replication and the inferred Rep motifs necessary for endonuclease and helicase activity

Accession number	Plant genomovirus	Collection year	Country	Plant species	Plant family	Nonanucleotide	Motif I	Motif II	GRS	Motif III	Walker A	Walker B	Motif C
Gemyctrularvirus													
MH939384	PaGmV 11*	2017	US	Solanaceae sp.	Solanaceae	TAATGTTAT	LLTYAQ	IHLHV	DVFDVGGCHPNVKCG	YAIK	GESRLGKT	VFDDM	WLAN
MH939425	PaGmV 12	2017	US	<i>Larrea tridentata</i>	Zygophyllaceae	TAATGTTAT	LLTYAQ	THLHV	NFFDVGHHPNITPSR	YAVK	GPSRLGKT	VFDDM	WLAN
MH939427	PaGmV 13*	2012	ZA	<i>Willdenowia</i> sp.	Restionaceae	TAATGTTAT	LLTYSQ	THLHV	DYFDVQGGHHPNIVPSR	YATK	GASRLGKT	VFDDM	WLAN
MH939434	PaGmV 13*	2012	ZA	<i>Asparagus declinatus</i>	Asparagaceae	TAACACTGT	LLTYSQ	THLHV	DYFDVQGGHHPNIVPSR	YATK	EASDLMPT	VFDDM	WLAN
MH939452	PaGmV 14*	2012	ZA	<i>Hordeum murinum</i>	Poaceae	TAACACTGT	LLTYAQ	THLHV	DVFDVGGCHPNIVPSR	YAVK	GASRLGKT	IFDDM	WLAN
MH939377	PaGmV 15*	2008	BR	<i>Phaseolus lunatus</i>	Leguminosae	TAACACTGT	LLTYSQ	IHLHC	DIFDVGGRHPNIVGSY	YAIK	GNSTRGKT	IFDDI	WLSN
MH939378	PaGmV 15*	2008	BR	<i>Ipomoea batatas</i>	Convolvulaceae	TAATATTAT	LLTYSQ	IHLHC	DIFDVGGRHPNIVGSY	YAIK	GNSTRGKT	IFDDI	WLSN
MH939379	PaGmV 15*	2008	BR	<i>Phaseolus vulgaris</i>	Leguminosae	TAATATTAT	LLTYSQ	IHLHC	DIFDVGGRHPNIVGSY	YAIK	GNSTRGKT	IFDDI	WLSN
MH939380	PaGmV 15*	2008	BR	<i>Ipomoea alba</i>	Convolvulaceae	TAATATTAT	LLTYAQ	LHLHV	DIFDVGGRHPNIEQSY	YAIK	GNSTRGKT	IFDDI	WLSN
MH939396	PaGmV 16*	2012	ZA	<i>Phalaris minor</i>	Poaceae	TAATATTAT	LLTYSQ	LHLHV	DILDVDGRHANVEPSA	YAIK	GGTRTGKT	VFDDI	WVCN
MH939397	PaGmV 17*	2012	ZA	<i>Sarcocornia perennis</i>	Amaranthaceae	TAATGTTAT	LVTYAQ	LHLHV	DILDVDGRHPNLAPIK	YAIK	GRSRTGKT	VFDDI	WVCN
MH939432	PaGmV 18*	2012	ZA	<i>Avena sativa</i>	Poaceae	TAATGTTAT	LLTYAQ	VHLHC	DVFDVEGRHPNISPFSK	YAIK	GESRTGKT	VFDDL	YLAN
MH939447	PaGmV 18*	2012	ZA	<i>Avena sativa</i>	Poaceae	TAATGTTAT	LLTYAQ	VHLHC	DVFDVEGRHPNISPFSK	YAIK	GESITGKT	VFDDL	YLAN
MH939448	PaGmV 18*	2012	ZA	<i>Avena sativa</i>	Poaceae	TAATATTAT	LLTYAQ	VHLHC	DVFDVEGRHPNISPFSK	YAIK	GESRTGKT	VFDDL	YLAN
MH939451	PaGmV 18*	2012	ZA	<i>Raphanus</i> sp.	Brassicaceae	TAAACATTGT	LLTYAQ	VHLHC	DVFDVEGRHPNISPFSK	YAIK	GESRTGKT	VFDDL	YLAN
MH939446	PaGmV 19*	2012	ZA	<i>Dischisma capitatum</i>	Scrophulariaceae	TAATGTTAT	LLTYPQ	IHLHC	RAFDVEGHHPNVSPFSK	YATK	GQSRGKT	IFDDI	FGCQ
MH939431	PaGmV 20*	2012	ZA	<i>Lebeckia</i> sp.	Leguminosae	TAATATTAT	LLTYSQ	IHFHA	RDWIDIEGTHPNIKKGG	YACK	GESRLGKT	VFDDI	WVAN
MH939436	PaGmV 21*	2012	ZA	<i>Asparagus</i> sp.	Asparagaceae	TAATGTTAT	LFTYAH	VHYHV	DCFVDVAGFHPNIVASR	YAIK	GDTKLGKT	VFDDI	WVAN
MH939442	PaGmV 22*	2012	ZA	<i>Hypochoeris</i> sp.	Asteraceae	TAATGTTAT	LVTYSH	IHYHV	DIFDVHGFHPNINPSR	YVIK	GPTRTGKT	IFDDI	WVSN
MH939445	PaGmV 22*	2012	ZA	<i>Tribolium uniolae</i>	Poaceae	TAATGTTAT	LVTYSH	IHYHV	DIFDVHGFHPNINPSR	YVIK	GPTRTGKT	IFDDI	WVSN
MK947372	PaGmV 25*	2018	FR	<i>Opuntia</i> sp.	Cactaceae	TAATGTTAT	LITYAQ	IHLHA	TVFDVGGCHPNIEKSR	YAIK	GGSQLGKT	IMDDI	WCCN

Table 1 (continued)

Accession number	Plant genomovirus	Collection year	Country	Plant species	Plant family	Nonanucleotide	Motif I	Motif II	GRS	Motif III	Walker A	Walker B	Motif C
Gemyduguvirus													
MH939370	PaGmV 4*	2015	BR	<i>Macropitium</i> sp.	Leguminosae	TAAATGTTAT	LLTYPQ	THYHA	RIFDVGHPNILLRGR	YATK	GPSRTGKT	VFDDI	WCNN
MH939417	PaGmV 5*	2008	BR	Apocynaceae sp.	Apocynaceae	TAAATGTTAT	LLTYAQ	THYHA	RIFDIDTYHPNILLRGR	YATK	GPSRTGKT	VFDDI	WCNN
MH939418	PaGmV 5*	2008	BR	<i>Vigna</i> sp.	Leguminosae	TAAATGTTAT	LLTYAQ	THYHA	RIFDIDTYHPNILLRGR	YATK	GPSRTGKT	VFDDI	WCNN
MH939419	PaGmV 5*	2008	BR	<i>Merremia aegyptia</i>	Convolvulaceae	TAAATATTAT	LLTYAQ	THYHA	RIFDIDTYHPNILLRGR	YATK	GPSRTGKT	VFDDI	WCNN
MH939420	PaGmV 5*	2008	BR	<i>Phaseolus lunatus</i>	Leguminosae	TAAATATTAT	LLTYAQ	THYHA	RIFDIDTYHPNILLRGR	YATK	GPSRTGKT	VFDDI	WCNN
MH939421	PaGmV 5*	2008	BR	<i>Conyza bonariensis</i>	Asteraceae	TAAATATTAT	LLTYAQ	THYHA	RIFDIDTYHPNILLRGR	YATK	GPSRTGKT	VFDDI	WCNN
MH939422	PaGmV 5*	2008	BR	Apocynaceae sp.	Apocynaceae	TAAATATTAT	LLTYAQ	THYHA	RIFDIDTYHPNILLRGR	YATK	GPSRTGKT	VFDDI	WCNN
MH939423	PaGmV 5*	2008	BR	Apocynaceae sp.	Apocynaceae	TAAATATTAT	LLTYAQ	THYHA	RIFDIDTYHPNILLRGR	YATK	GPSRTGKT	VFDDI	WCNN
MH939424	PaGmV 5*	2008	BR	Apocynaceae sp.	Apocynaceae	TAAATATTAT	LLTYAQ	THYHA	RIFDIDTYHPNILLRGR	YATK	GPSRTGKT	VFDDI	WCNN
Gemygorvirus													
MH939361	PaGmV 1*	2015	BR	Poaceae sp.	Poaceae	TAAATGTTAT	LFTYAQ	THLHA	RRFDVGFHPNVSFSR	YAIK	GPTRTGKT	IFDDM	WLSN
MH939362	PaGmV 1*	2015	BR	<i>Stylosanthes</i> sp.	Leguminosae	TAAATGTTAT	LFTYAQ	THLHA	RRFDVGFHPNVSFSR	YAIK	GPTRTGKT	IFDDM	WLSN
MH939367	PaGmV 1*	2015	BR	Poaceae sp.	Poaceae	TAAATGTTAT	LFTYAQ	THLHA	RRFDVGFHPNVSFSR	YAIK	GPTRTGKT	IFDDM	WLSN
MH939368	PaGmV 1*	2015	BR	<i>Stylosanthes</i> sp.	Leguminosae	TAAATGTTAT	LFTYAQ	THLHA	RRFDVGFHPNVSFSR	YAIK	GPTRTGKT	IFDDM	WLSN
MH939371	PaGmV 1*	2015	BR	<i>Manihot</i> sp.	Euphorbiaceae	TAAATGTTAT	LFTYAQ	THLHA	RRFDVGFHPNVSFSR	YAIK	GPTRTGKT	VFDDM	WLSN
MH939372	PaGmV 1*	2015	BR	<i>Centratherum punctatum</i>	Asteraceae	TAAATGTTAT	LFTYAQ	THLHA	RRFDVGFHPNVSFSR	YAIK	GPTRTGKT	VFDDM	WLSN
MH939373	PaGmV 1*	2015	BR	<i>Solanum lycopersicum</i>	Solanaceae	TAAATGTTAT	LFTYAQ	THLHA	RRFDVGFHPNVSFSR	YAIK	GPTRTGKT	VFDDM	WLSN
MK947373	PaGmV 26*	2018	FR	<i>Opuntia</i> sp.	Cactaceae	TAAATGTTAT	LLTYSQ	IHLHA	RFADVEGRHPNVQPFQ	YAIK	GPTRLGKT	VFDDM	WLCN
Gemykibivirus													
MH939363	PaGmV 2*	2015	BR	Poaceae sp.	Poaceae	TAAATGTTAT	LLTYAQ	VHLHA	RVFDVAGRHPNVVAGY	YATK	GDTRLGKT	VFDDM	YCCN
MH939364	PaGmV 2*	2015	BR	<i>Sida</i> sp.	Malvaceae	TAAATGTTAT	LLTYAQ	VHLHA	RVFDVAGRHPNVVAGY	YATK	GDTRLGKT	VFDDM	YCCN
MH939365	PaGmV 2*	2015	BR	<i>Mimosa</i> sp.	Leguminosae	TAAATGTTAT	LLTYAQ	VHLHA	RVFDVAGRHPNVVAGY	YATK	GDTRLGKT	VFDDM	YCCN
MH939366	PaGmV 2*	2015	BR	<i>Sida</i> sp.	Malvaceae	TAAATGTTAT	LLTYAQ	VHLHA	RVFDVAGRHPNVVAGY	YATK	GDTRLGKT	VFDDM	YCCN
MH939411	PaGmV 2*	2015	BR	<i>Sida</i> sp.	Malvaceae	TAAATGTTAT	LLTYAQ	VHLHA	RVFDVAGRHPNVVAGY	YATK	GDTRLGKT	VFDDM	YCCN

Table 1 (continued)

Accession number	Plant genomovirus	Collection year	Country	Plant species	Plant family	Nonanucleotide	Motif I	Motif II	GRS	Motif III	Walker A	Walker B	Motif C
MH939412	PaGmV 2*	2015	BR	<i>Adenocalymma</i> sp.	Bignoniaceae	TAAATGTTAT	LLTYAQ	VHLHA	RVFDVAGRHPNVVAGY	YATK	GDTRLGKT	VFDDM	YCCN
MH939413	PaGmV 2*	2015	BR	<i>Sida</i> sp.	Malvaceae	TAAATGTTAT	LLTYAQ	VHLHA	RVFDVAGRHPNVVAGY	YATK	GDTRLGKT	VFDDM	YCCN
MH939414	PaGmV 2*	2015	BR	<i>Ipomoea</i> sp.	Convolvulaceae	TAAATGTTAT	LLTYAQ	VHLHA	RVFDVAGRHPNVVAGY	YATK	GDTRLGKT	VFDDM	YCCN
MH939415	PaGmV 2*	2015	BR	<i>Sida</i> sp.	Malvaceae	TAAATGTTAT	LLTYAQ	VHLHA	RVFDVAGRHPNVVAGY	YATK	GDTRLGKT	VFDDM	YCCN
MH939416	PaGmV 2*	2015	BR	<i>Bidens pilosa</i>	Asteraceae	TAAATATTAT	LLTYAQ	VHLHA	RVFDVAGRHPNVVAGY	YATK	GDTRLGKT	VFDDM	YCCN
MK947374	PaGmV 27*	2015	US	<i>Cylindropuntia ramosissima</i>	Cactaceae	TAAATGTTAT	LLTYPQ	VHLHA	RIFDVGDMHPNVVRGY	YAIK	GPTRLGKT	IFDDM	YISN
MK947375	PaGmV 28*	2017	US	<i>Carnegiea gigantea</i>	Cactaceae	TAAACACTGT	LLTYPQ	LHLHA	RVFDVDGKHPNVVRGY	YACK	GGTRLGKT	IFDDM	YIAN
MK947376	PaGmV 29*	2017	US	<i>Carnegiea gigantea</i>	Cactaceae	TAAATGTTAT	LEFTYAQ	THLHA	DIFDVGDFHPNVVPSR	YATK	GPSRKGKT	VFDDM	YLAN
MH939438	PaGmV 3*	2012	ZA	<i>Cynodon</i> sp.	Poaceae	TAAATGTTAT	LLTYPQ	VHLHA	RVFDVDGHPNVSRGY	YAIK	GTTRLGKT	IFDDV	WLAN
MH939439	PaGmV 3*	2012	ZA	<i>Hypochoeris radicata</i>	Asteraceae	TAAATGTTAT	LLTYPQ	VHLHA	RVFDVDGHPNVSRGY	YAIK	GTTRLGKT	IFDDV	WLAN
MH939369	PaGmV 6	2015	BR	<i>Manihot</i> sp.	Euphorbiaceae	TAAATATTAT	LLTYSQ	THLHA	SIFDVDNHHPNVSATH	YACK	GPSRMGKT	IFDDL	WISN
MH939453	PaGmV 6	2012	ZA	<i>Vicia faba</i>	Leguminosae	TAAATATTAT	LLTYSQ	THLHA	SVFDVNDCHPNVSATH	YACK	GPSRMGKT	IFDDI	WISN
Gemykolovirus													
MH939386	PaGmV 10	2015	AU	<i>Solanum mauritianum</i>	Solanaceae	TAAATGTTAT	MLTYPT	PHLHA	ATFKIGTRVPNIRVRR	YVGK	GPTRTGT	IFDDM	FICN
MH939387	PaGmV 10	2015	AU	<i>Solanum mauritianum</i>	Solanaceae	TAAAAATTA	MLTYPT	PHLHA	ATFKIGTRVPNIRVRR	YVGK	GPTRTGT	IFDDM	FICN
MH939388	PaGmV 10	2015	AU	<i>Solanum mauritianum</i>	Solanaceae	TAAATATTAT	MLTYPT	PHLHA	ATFKIGTRVPNIRVRR	YVGK	GPTRTGT	IFDDM	FICN
MH939374	PaGmV 7*	2015	BR	<i>Poaceae</i> sp.	Poaceae	TAAATGTTAT	LLTYSQ	PHLHA	SLFFVGGRRANIRRFS	YVGK	GATRLGKT	IFDDI	WLCN
MH939375	PaGmV 7*	2015	BR	<i>Ipomoea</i> sp.	Convolvulaceae	TAAATGTTAT	LLTYSQ	PHLHA	SLFFVGGRRANIRRFS	YVGK	GATRLGKT	VFDDI	WLCN
MH939376	PaGmV 7*	2015	BR	<i>Herissantia</i> sp.	Malvaceae	TAAACTACTAT	LLTYSQ	PHLHA	SLFFVGGRRANIRRFS	YVGK	GATRLGKT	VFDDI	WLCN
MH939410	PaGmV 7*	2015	BR	<i>Passiflora edulis</i>	Passifloraceae	TAAATGTTAT	LLTYSQ	PHLHA	SLFFVGGRRANIRRFS	YVGK	GATRLGKT	IFDDI	WLCN
MH939381	PaGmV 8*	2012	ZA	<i>Bromus diandrus</i>	Poaceae	TAAATGTTAT	MLTYPT	PHLHA	RTFTVTGTRIPNIRVRR	YVGK	GATRLGKT	IWDDM	FICN

Table 1 (continued)

Accession number	Plant genomovirus	Collection year	Country	Plant species	Plant family	Nonanucleotide	Motif I	Motif II	GRS	Motif III	Walker A	Walker B	Motif C
MH939389	PaGmV 8*	2012	ZA	<i>Avena byzantina</i>	Poaceae	TAAATGTTAT	MLTYPT	PEFHA	RTFTVGTTRVPNIRVRR	YVGK	GATRLGKT	IWDDM	FICN
MH939391	PaGmV 8*	2012	ZA	<i>Helichrysum revolutum</i>	Asteraceae	TAAAGATTG	MLTYPT	PEFHA	RTFTVGTTRIPNIRVRR	YVGK	GATRLGKT	IWDDM	FICN
MH939393	PaGmV 8*	2012	ZA	<i>Avena byzantina</i>	Poaceae	TAAATGTTAT	MLTYPT	PEFHA	RTFTVGTTRIPNIRVRR	YVGK	GATRLGKT	IWDDM	FICN
MH939395	PaGmV 8*	2012	ZA	<i>Avena byzantina</i>	Poaceae	TAAATGTTAT	MLTYPT	PEFHA	RTFTVGTTRIPNIRVRR	YVGK	GATRLGKT	IWDDM	FICN
MH939399	PaGmV 8*	2012	ZA	<i>Avena fatua</i>	Poaceae	TAAATACTAT	MLTYPT	PEFHA	RTFTVGTTRVPNIRVRR	YVGK	GATRLGKT	IWDDM	FICN
MH939400	PaGmV 8*	2012	ZA	<i>Asparagus rubicundus</i>	Asparagaceae	TAAACATTGT	MLTYPT	PEFHA	RTFTVGTTRIPNIRVRR	YVGK	GATRLGKT	IWDDM	FICN
MH939403	PaGmV 8*	2012	ZA	<i>Sarcocornia perennis</i>	Amaranthaceae	TAAATATTAT	MLTYPT	PEFHA	RTFTVGTTRIPNIRVRR	YVGK	GATRLGKT	IWDDM	FICN
MH939404	PaGmV 8*	2012	ZA	<i>Bromus diandrus</i>	Poaceae	TAAATGTTAT	MLTYPT	PEFHA	RTFTVGTTRIPNIRVRR	YVGK	GATRLGKT	IWDDM	FICN
MH939405	PaGmV 8*	2012	ZA	<i>Manulea altissima</i>	Scrophulariaceae	TAAATGTTAT	MLTYPT	PEFHA	RTFTVGTTRIPNIRVRR	YVGK	GATRLGKT	IWDDM	FICN
MH939433	PaGmV 8*	2012	ZA	<i>Asparagus declinatus</i>	Asparagaceae	TAAATGTTAT	MLTYPT	PEFHA	RTFTVGTTRIPNIRVRR	YVGK	GAIWLGKT	IWDDM	FICN
MH939441	PaGmV 8*	2012	ZA	<i>Emex australis</i>	Polygonaceae	TAAACATTGT	MLTYPT	PEFHA	RTFTVGTTRIPNIRVRR	YVGK	GAIWLGKT	IWDDM	FICN
MH939443	PaGmV 8*	2012	ZA	Poaceae sp.	Poaceae	TAAATGTTAT	MLTYPT	PEFHA	RTFTVGTTRIPNIRVRR	YVGK	GATRLGKT	IWDDM	FICN
MH939382	PaGmV 9*	2012	ZA	<i>Avena byzantina</i>	Poaceae	TAAATATTAT	MLTYPT	PEFHA	RTFTVGTTRVPNIRVRR	YVGK	GATRLGKT	VFDDM	FICN
MH939383	PaGmV 9*	2012	ZA	<i>Bromus diandrus</i>	Poaceae	TAAATGTTAT	MLTYPT	PEFHA	RTFTVGTTRVPNIRVRR	YVGK	GATRLGKT	VFDDM	FICN
MH939390	PaGmV 9*	2012	ZA	<i>Cyclopia genistoides</i>	Leguminosae	TAAACATTGT	MLTYPT	PEFHA	RTFTVGTTRVPNIRVRR	YVGK	GATRLGKT	VFDDM	FICN
MH939392	PaGmV 9*	2012	ZA	<i>Lolium perenne</i>	Poaceae	TAAATATTAT	MLTYPT	PEFHA	RTFTVGTTRVPNIRVRR	YVGK	GATRLGKT	VFDDM	FICN
MH939394	PaGmV 9*	2012	ZA	<i>Avena byzantina</i>	Poaceae	TAAATGTTAT	MLTYPT	PEFHA	RTFTVGTTRVPNIRVRR	YVGK	GATRLGKT	VFDDM	FICN
MH939398	PaGmV 9*	2012	ZA	<i>Avena byzantina</i>	Poaceae	TAAATGTTAT	MLTYPT	PEFHA	RTFTVGTTRVPNIRVRR	YVGK	GATRLGKT	VFDDM	FICN
MH939401	PaGmV 9*	2012	ZA	<i>Bromus diandrus</i>	Poaceae	TAAATGTTAT	MLTYPT	PEFHA	RTFTVGTTRVPNIRVRR	YVGK	GATRLGKT	VFDDM	FICN
MH939402	PaGmV 9*	2012	ZA	<i>Avena byzantina</i>	Poaceae	TAAACATTGT	MLTYPT	PEFHA	RTFTVGTTRVPNIRVRR	YVGK	GATRLGKT	VFDDM	FICN
MH939406	PaGmV 9*	2012	ZA	<i>Avena byzantina</i>	Poaceae	TAAACCAGG	MLTYPT	PEFHA	RTFTVGTTRVPNIRVRR	YVGK	GATRLGKT	VFDDM	FICN

Table 1 (continued)

Accession number	Plant genomovirus	Collection year	Country	Plant species	Plant family	Nonanucleotide	Motif I	Motif II	GRS	Motif III	Walker A	Walker B	Motif C
MH939407	PaGmV 9*	2012	ZA	<i>Bromus diandrus</i>	Poaceae	TAAATGTTAT	MLTYPT	PEFHA	RTFTVGTTRVVPNIRVRR	YVGK	GATRLGKT	VFDDM	FICN
MH939408	PaGmV 9*	2012	ZA	<i>Avena byzantina</i>	Poaceae	TAAATATTAT	MLTYPT	PEFHA	RTFTVGTTRVVPNIRVRR	YVGK	GATRLGKT	VFDDM	FICN
MH939409	PaGmV 9*	2012	ZA	<i>Avena byzantina</i>	Poaceae	TAAAGGTTCA	MLTYPT	PEFHA	RTFTVGTTRVVPNIRVRR	YVGK	GATRLGKT	VFDDM	FICN
MH939426	PaGmV 9*	2012	ZA	Poaceae sp.	Poaceae	TAAATATTAT	MLTYPT	PEFHA	PTFTVGTTRVVPNLRVRL	YVAK	GATRLGKT	VFDDM	FICN
MH939428	PaGmV 9*	2012	ZA	<i>Avena sativa</i>	Poaceae	TAAATGTTAT	MLTYPT	PEFHA	RTFTVGTTRVVPNIRVRR	YVGK	GATRLGKT	VFDDM	FICN
MH939429	PaGmV 9*	2012	ZA	<i>Avena sativa</i>	Poaceae	TAAATGTTAT	MLTYPT	PEFHA	RTFTIGTRVVPNIRVRR	YVGK	GATRLGKT	VFDDM	FICN
MH939430	PaGmV 9*	2012	ZA	<i>Avena sativa</i>	Poaceae	TAAATATTAT	MLTYPT	PEFHA	RTFTVGTTRVVPNIRVRR	YVGK	GATRLGKT	VFDDM	FICN
MH939437	PaGmV 9*	2012	ZA	<i>Hordeum vulgare</i>	Poaceae	TAAATGTTAT	MLTYPT	PEFHA	RTFTVGTTRVVPNIRVRR	YVGK	GATRLGKT	VFDDM	FICN
Genykronezavirus													
MH939385	PaGmV 23*	2017	US	Solanaceae sp.	Solanaceae	TAAACTAT	FLYSQ	HHYHV	RTFDVGGCHPNFKSVR	YCLK	GRSRLGKT	VMDDI	WCSN
MH939440	PaGmV 23*	2012	ZA	<i>Ehrlharta longiflora</i>	Poaceae	TAAATGTTAT	FLYSQ	HHYHV	RVFDVGGCHPNFKSVR	YCLK	GNSRLGKT	VMDDI	WCSN
MH939449	PaGmV 23*	2012	ZA	<i>Raphanus</i> sp.	Brassicaceae	TAAATGTTAT	FLYSQ	HHYHV	RVFDVGGCHPNFKSVR	YCLK	GNSRLGKT	VMDDI	WCSN
MH939450	PaGmV 23*	2012	ZA	<i>Raphanus</i> sp.	Brassicaceae	TAAATGTTAT	PVCIPL	POFYF	RVFDVGGCHPNFKSVR	YCLK	GNSRLGKT	VMDDI	WCSN
MH939435	PaGmV 24*	2012	ZA	Poaceae sp.	Poaceae	TAAATGTTAT	FLYSQ	HHYHV	RTFDVGGCHPNFKSVR	YCLK	GQSRLGKT	VMDDI	WCTN
MH939444	PaGmV 24*	2012	ZA	<i>Salvia africana-coerulea</i>	Lamiaceae	TAAACTAT	FLYSQ	HHYHV	RTFDVGGCHPNFKSVR	YCLK	GQSRLGKT	VMDDI	WCTN

*members of new species based on the current demarcation threshold for the family Genomoviridae

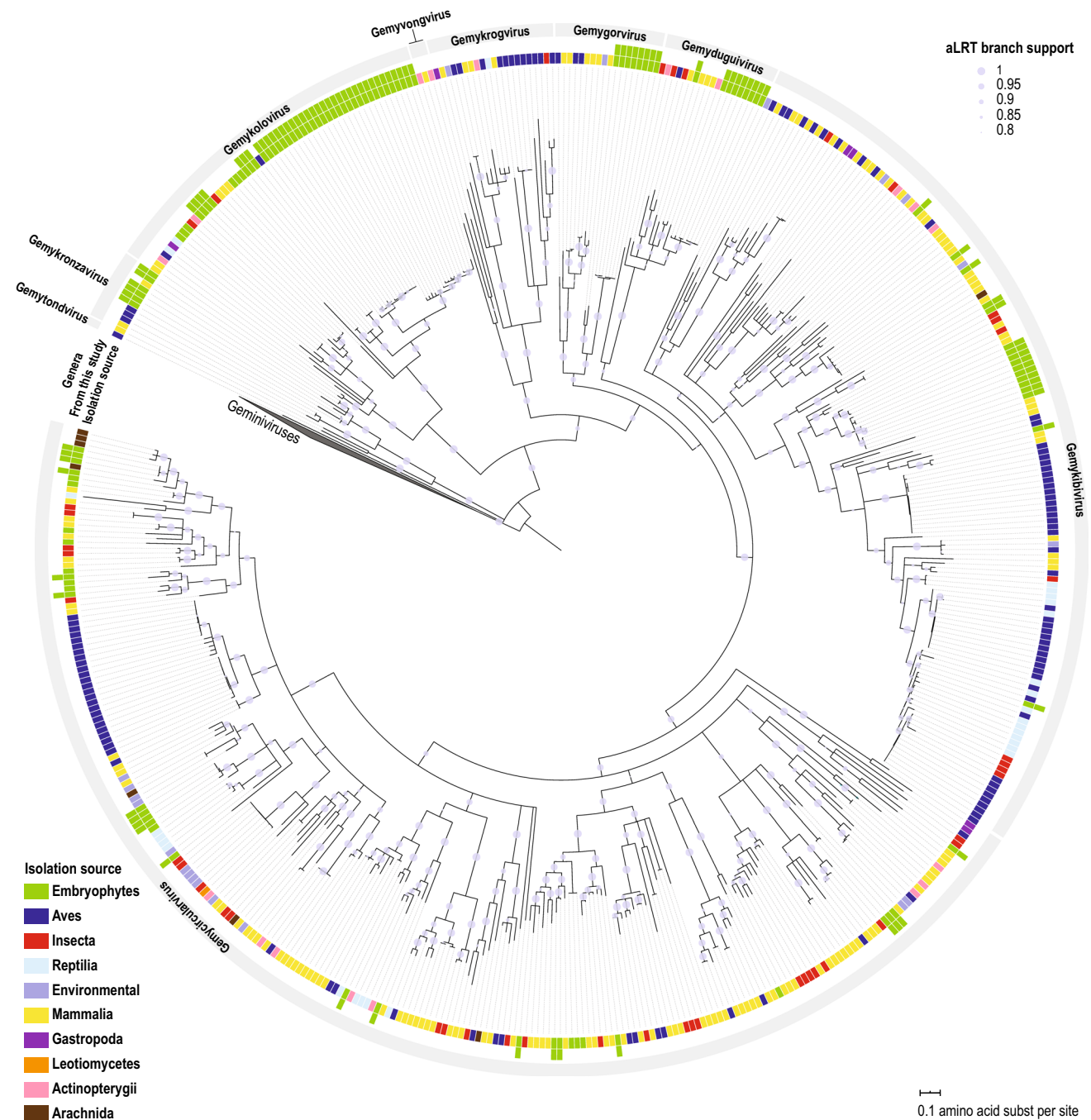


Fig. 1 Maximum-likelihood phylogenetic tree of the Rep amino acid sequences encoded by the geminiviruses from this study in relation to those of other geminiviruses available in the GenBank database (downloaded 1 Feb 2020). The tree was rooted with Rep amino acid sequences from the family *Geminiviridae*. Branches with < 0.8 aLRT

support have been collapsed. The tree is further demarcated by the source from which sequences were obtained: Actinopterygii ($n = 18$), Arachnida ($n = 13$), Aves ($n = 113$), Embryophyte ($n = 116$); environmental ($n = 20$), Gastropoda ($n = 6$), Insecta ($n = 44$), Leotiomyces ($n = 1$), Mammalia ($n = 155$), and Reptilia ($n = 28$)

genus *Gemykroznavirus* (Fig. 1). The other PaGmVs were tentatively assigned to two of the larger geminivirus genera with PaGmV 11-22, PaGmV 25 ($n = 21$) to the genus *Gemyrcircularvirus*, and PaGmV 2-3 and 27-29 ($n = 17$) to

the genus *Gemykibivirus*. The phylogenetic analysis showed that there is no clear correlation between the geminivirus genera and the sample type from which the sequences were obtained (Fig. 1 and Fig. 2B).

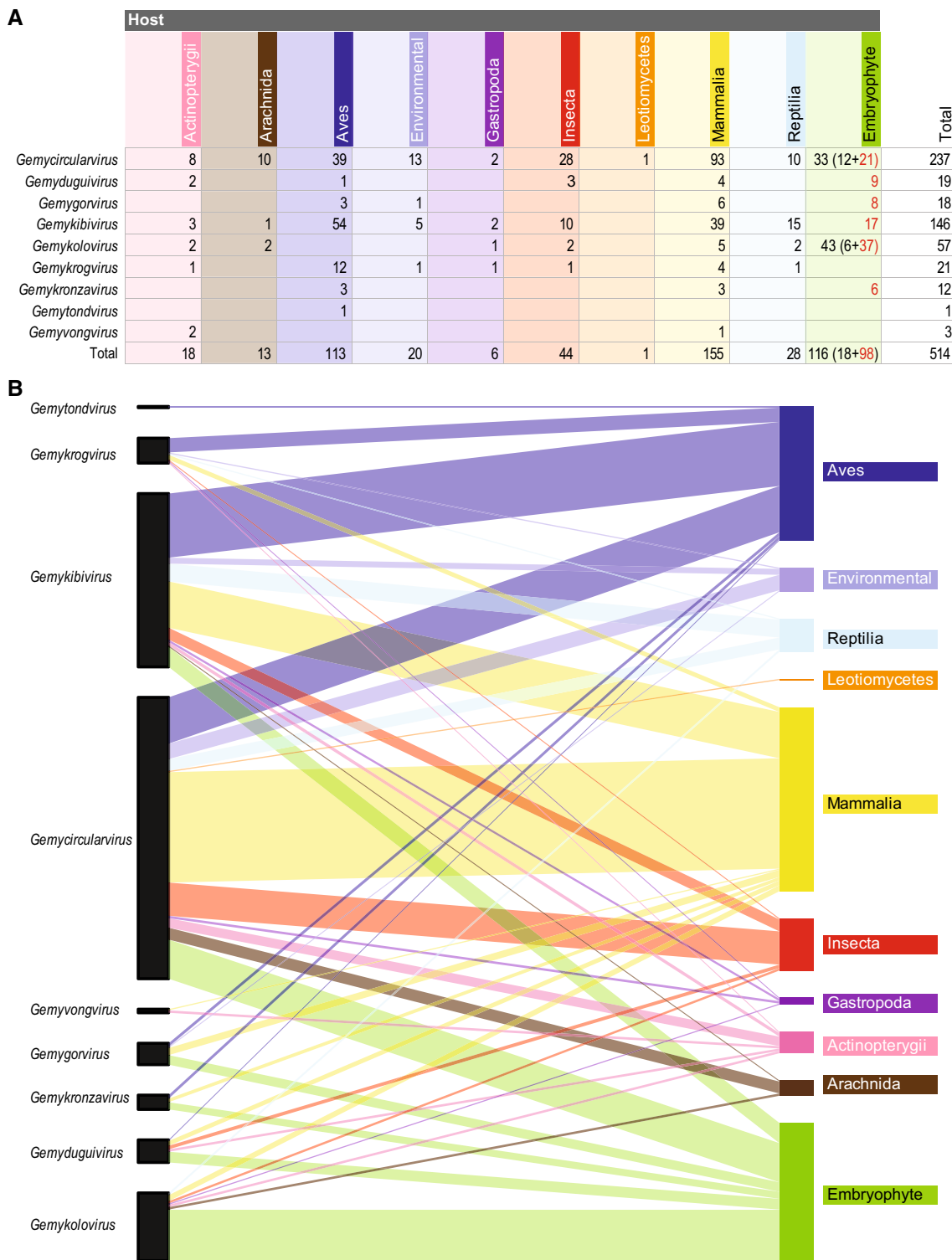


Fig. 2 **A.** Summary table of the number of viruses identified in this study (highlighted in red font) in relation to those from other sample sources and their genus assignments. **B.** Bipartite plot of the host/virus species associations inferred using R with the bipartite graph package [8]

The inferred Rep proteins encoded by the PaGmVs all contain the conserved motifs necessary for endonuclease and helicase activity (Table 1) that are present in most

other evolutionarily related Rep proteins. The genome sequences of the PaGmVs share >63.3% pairwise identity with other genomoviruses. The Rep amino acid sequence

of the PaGmVs share >63.7% identity with those of other genomoviruses, while those of their CPs share >37.9% identity (Supplementary Table 1).

This study further highlights the broad diversity of genomoviruses that are associated with plants. Before this study, no viruses in the genera *Gemykroznavirus*, *Gemygorvirus*, *Gemyduguivirus* or *Gemykibivirus* had been found associated with plants. Our research has identified six species of *Gemykroznavirus*, eight species of *Gemygorvirus*, nine species of *Gemyduguivirus*, and 17 species of *Gemykibivirus* associated with plants (Fig. 2). Additionally, the 37 gemykoloviruses identified here demonstrate that plant-associated viruses are the predominant members of this genus (Fig. 2). Although there are no confirmed hosts for any of the genomoviruses discovered here, it is likely that they infect fungi that are associated with the plants from which they were obtained, but it is also possible that they infect both fungi and their plants hosts. It is probable that additional diverse plant-associated genomoviruses will be identified, especially given the close association between plants and fungi. Furthermore, it is also likely that multipartite genomo-like viruses will be identified that are similar to the tripartite ssDNA virus recently reported infecting *Fusarium graminearum*, whose Rep shares 24–33% amino acid sequence identity with genomovirus Reps [17].

Acknowledgements This research was partially funded by grants from Embrapa, CNPq and FAPDF (Brazil). We thank CNPq for the fellowships to NSL and ALML, and to CAPES for the scholarships to RAA and GBB. This research was partially funded by a Student and Early Career Grant awarded to RSF by the Next Generation Sonoran Desert Researchers (N-Gen).

Compliance with ethical standards

Conflict of interest The authors declare there are no conflicts of interest.

Research involving human participants and/or animals The research did not involve human participants or animals.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Bernardo P, Charles-Dominique T, Barakat M, Ortet P, Fernandez E, Filloux D, Hartnady P, Rebelo TA, Cousins SR, Mesleard F, Cohez D, Yavercovski N, Varsani A, Harkins GW, Peterschmitt M, Malmstrom CM, Martin DP, Roumagnac P (2018) Geometagenomics illuminates the impact of agriculture on the distribution and prevalence of plant viruses at the ecosystem scale. *ISME J* 12:173–184
- Chabi-Jesus C, Najar A, Fontenele RS, Kumari SG, Ramos-Gonzalez PL, Freitas-Astua J, Kraberger S, Varsani A (2020) Viruses representing two new genomovirus species identified in citrus from Tunisia. *Arch Virol*:1–5
- Darriba D, Taboada GL, Doallo R, Posada D (2011) ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27:1164–1165
- Dayaram A, Opong A, Jaschke A, Hadfield J, Baschiera M, Dobson RC, Offei SK, Shepherd DN, Martin DP, Varsani A (2012) Molecular characterisation of a novel cassava associated circular ssDNA virus. *Virus Res* 166:130–135
- de Rezende RR, Mar TB, Paez LMC, Silva Xavier AD, Xavier CAD, Navas-Castillo J, Zerbini FM, Alfenas-Zerbini P (2018) Complete genome sequences of two gemycircularviruses associated with non-cultivated plants in Brazil. *Arch Virol* 163:3163–3166
- Dellaporta SL, Wood J, Hicks JB (1983) A plant DNA miniprep: version II. *Plant Mol Biol Rep* 1:19–21
- Dormann CF, Fründ J, Blüthgen N, Gruber B (2009) Indices, graphs and null models: analyzing bipartite ecological networks. *Open Ecol J* 2
- Du Z, Tang Y, Zhang S, She X, Lan G, Varsani A, He Z (2014) Identification and molecular characterization of a single-stranded circular DNA virus with similarities to *Sclerotinia sclerotiorum* hypovirulence-associated DNA virus 1. *Arch Virol* 159:1527–1531
- Filloux D, Dallot S, Delaunay A, Galzi S, Jacquot E, Roumagnac P (2015) Metagenomics approaches based on virion-associated nucleic acids (VANA): an innovative tool for assessing without a priori viral diversity of plants. *Methods Mol Biol* 1302:249–257
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59:307–321
- Jackman SD, Vandervalk BP, Mohamadi H, Chu J, Yeo S, Hammond SA, Jahesh G, Khan H, Coombe L, Warren RL, Birol I (2017) ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. *Genome Res* 27:768–777
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649
- Kraberger S, Farkas K, Bernardo P, Booker C, Arguello-Astorga GR, Mesleard F, Martin DP, Roumagnac P, Varsani A (2015) Identification of novel Bromus- and Trifolium-associated circular DNA viruses. *Arch Virol* 160:1303–1311
- Krupovic M, Ghabrial SA, Jiang D, Varsani A (2016) Genomoviridae: a new family of widespread single-stranded DNA viruses. *Arch Virol* 161:2633–2643
- Lamas NS, Fontenele RS, Melo FL, Costa AF, Varsani A, Ribeiro SG (2016) Complete Genome Sequence of a Genomovirus Associated with Common Bean Plant Leaves in Brazil. *Genome Announc* 4
- Li P, Wang S, Zhang L, Qiu D, Zhou X, Guo L (2020) A tripartite ssDNA mycovirus from a plant pathogenic fungus is infectious as cloned DNA and purified virions. *Sci Adv* 6:eaay9634
- Male MF, Kami V, Kraberger S, Varsani A (2015) Genome Sequences of Poaceae-Associated Gemycircularviruses from the Pacific Ocean Island of Tonga. *Genome Announc* 3
- Marzano SL, Domier LL (2016) Novel mycoviruses discovered from metatranscriptomics survey of soybean phyllosphere phyto-biomes. *Virus Res* 213:332–342
- Muhire BM, Varsani A, Martin DP (2014) SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. *Plos One* 9:e108277
- Stover BC, Muller KF (2010) TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinform* 11:7

22. Varsani A, Krupovic M (2017) Sequence-based taxonomic framework for the classification of uncultured single-stranded DNA viruses of the family Genomoviridae. *Virus Evol* 3:vew037
23. Wren JD, Roossinck MJ, Nelson RS, Scheets K, Palmer MW, Melcher U (2006) Plant virus biodiversity and ecology. *PLoS Biol* 4:e80
24. Yu X, Li B, Fu Y, Jiang D, Ghabrial SA, Li G, Peng Y, Xie J, Cheng J, Huang J, Yi X (2010) A geminivirus-related DNA mycovirus that confers hypovirulence to a plant pathogenic fungus. *Proc Natl Acad Sci USA* 107:8387–8392
25. Zhang W, Olson NH, Baker TS, Faulkner L, Agbandje-McKenna M, Boulton MI, Davies JW, McKenna R (2001) Structure of the maize streak virus geminate particle. *Virology* 279:471–477

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.