Arch Virol
DOI 10.1007/s00705-017-3579-7

ANNOTATED SEQUENCE RECORD

Complete genome sequence of maize yellow striate virus, a new cytorhabdovirus infecting maize and wheat crops in Argentina

Fernanda Maurino[1,2] · Analía D. Dumón[1] · Gabriela Llauger[3] · Vanina Alemandri[1] · Luis A. de Haro[2,3] · M. Fernanda Mattio[1] · Mariana del Vas[2,3] · Irma Graciela Laguna[1,2] · María de la Paz Giménez Pecci[1]

Received: 18 April 2017 / Accepted: 14 September 2017

Abstract A rhabdovirus infecting maize and wheat crops in Argentina was molecularly characterized. Through next-generation sequencing (NGS) of symptomatic leaf samples, the complete genome was obtained of two isolates of maize yellow striate virus (MYSV), a putative new rhabdovirus, differing by only 0.4% at the nucleotide level. The MYSV genome consists of 12,654 nucleotides for maize and wheat virus isolates, and shares 71% nucleotide sequence identity with the complete genome of barley yellow striate mosaic virus (BYSMV, NC028244). Ten open reading frames (ORFs) were predicted in the MYSV genome from the antigenomic strand and were compared with their BYSMV counterparts. The highest amino acid sequence identity of the MYSV and BYSMV proteins was 80% between the L proteins, and the lowest was 37% between the proteins 4. Phylogenetic analysis suggested that the MYSV isolates are new members of the genus *Cytorhabdovirus*, family *Rhabdoviridae*. Yellow striate, affecting maize and wheat crops in Argentina, is an emergent disease that presents a potential economic risk for these widely distributed crops.

Handling Editor: Ralf Georg Dietzgen.

Fernanda Maurino and Analía D. Dumón contributed equally to this work.

✉ Fernanda Maurino
maurino.fernanda@inta.gob.ar

Analía D. Dumón
dumon.analia@inta.gob.ar

[1] Instituto de Patología Vegetal (IPAVE-CIAP-INTA), Av. 11 de Septiembre 4755 (X5020ICA), Córdoba, Argentina

[2] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Godoy Cruz 2290 (1425 FBQ) CABA, Buenos Aires, Argentina

[3] Instituto de Biotecnología (IB), CICVyA-INTA de los Reseros y Nicolás Repetto s/n (1686), Hurlingham, Buenos Aires, Argentina

Maize (*Zea mays* L.) and wheat (*Triticum aestivum* L.) are two of the world's main cereal crops, along with rice [1]. Argentina produced an average of 36.5 million tons of maize and 11.3 million tons of wheat in 2015/16 [1]. These species are affected by multiple viruses from several taxonomic groups worldwide, including rhabdoviruses. In 2012, a rhabdovirus infecting maize crops was reported in Argentina. Plants exhibiting shortened internodes, dwarfism, panicle sterility, and a mosaic of coarse and fine yellow striation on leaf blades and sheaths had been observed in several locations of the main maize production area in Argentina since 2000/01. Electron microscopy assays and sequencing of a conserved region of the polymerase gene demonstrated that the infection was caused by a tentatively new cytorhabdovirus, which was named "maize yellow striate virus" (MYSV) [2]. In 2015, a rhabdovirus isolate closely related to MYSV was reported affecting wheat crops. Wheat plants showing symptoms of mild chlorotic streaking on leaves, dwarfism, yellowing, and the presence of empty and deformed ears have been detected in several provinces of Argentina since 2007 [3].

In the present work, we describe the complete genome sequence of MYSV, confirm its occurrence in wheat plants, and propose to assign it to a tentative new species of the genus *Cytorhabdovirus*, family *Rhabdoviridae*. For this purpose, maize and wheat plants showing symptoms of yellow striation and dwarfism characteristic of MYSV infections were processed, and the viral genome was sequenced and analyzed.

Published online: 19 October 2017

Provenance of maize virus-infected plant material and determination of the genomic sequence. Leaf samples of five maize plants showing the previously described symptoms of MYSV were collected during 2014/15 from a field located in Sinsacate (Córdoba province, Argentina). For viral enrichment, 100 g of pooled samples were ground with four volumes of extraction buffer (0.1 M Tris-HCl, pH 8.4). The homogenate was filtered cold through four layers of cheesecloth. After centrifugation for 15 min at 8000 rpm and 4 °C in a JA14 rotor (Beckman Coulter, USA), the supernatant was filtered again and centrifuged for 1 h at 25,000 rpm in a SW28 rotor (Beckman). The pellet was resuspended in 300 µl of RNase-free water. Total RNA was extracted using an SV Total RNA Isolation System (Promega Corp., USA) following the manufacturer´s instructions. Total RNA was sent to INDEAR S.A. (Rosario, Argentina), where it was purified, sequenced on a WGS Illumina Hiseq 1500, and *de novo* assembled using A5 Pipeline software [4]. The identity of the individual scaffolds obtained was analyzed using BLASTn and BLASTx. One scaffold showed 71% nucleotide (nt) sequence identity to the complete genome of barley yellow striate mosaic virus (BYSMV, NC028244), with 87% coverage. The fragment of 12,638 nt was used as a reference sequence to map the cleaned reads using Geneious 10.0.5 (free trial) [5]. The MYSV genome obtained was analyzed to determine the quality of the assembly, using the Tablet 0.09.10.30 software [6].

Provenance of wheat virus-infected plant material and determination of the genomic sequence. The virus material was obtained from a single severely affected wheat plant (cv. Biointa 3005) collected in Río Cuarto (Córdoba, Argentina), maintained since 2013 in the laboratory by serial transmissions with the *Delphacodes kuscheli* vector. Total RNA was extracted from leaf tissue of this rhabdovirus-infected plant using TRIzol Reagent (Life Technologies, USA) following manufacturer's instructions. The extracted RNA was sent to INDEAR S.A. for NGS sequencing, where it was processed as described above for maize. One 12,652-nt scaffold shared 71% identity with the complete genome sequence of BYSMV (GenBank NC028244), with 87% coverage. The assembly was analyzed to determine its quality using Tablet 0.09.10.30 software.

To amplify and analyze the MYSV 5' and 3' ends, rapid amplification of cDNA ends (RACE) was performed [8] using 1 µg of total RNA from MYSV-infected wheat or maize leaves as template. For 5' RACE, the specific internal primer was 5' GTTAAAGTTCGAGATCACGA 3', nt 12353-12372, and the nested specific primer was 5' AGAGTATAGAGTGTTCGATTG 3', nt 12401-12421. For 3' RACE, the specific internal primer was 5' TAGCTGGCTATCATTAGGGC 3', nt 299-318, and the nested specific primer was 5' CGCCGGTACTTTATCAAACT 3' nt 260-279. The amplified products were purified using a QIAquick Gel Extraction Kit, (QIAGEN) and cloned into pGEM-T Easy Vector (Promega, USA). The complete MYSV genome sequence obtained from wheat was 12,654 nt long (GenBank accession number KY884672) with an average global coverage [7] of 1490.96X. From maize, the length was 12,654 nt with an average global coverage [7] of 466.10X (GenBank accession number KY884303).

The complete nucleotide sequence identity between the two MYSV genomes was 99.6%, suggesting that the two genome sequences correspond to different isolates of MYSV. The ORFs from the MYSV genome were identified using the Open Reading Frame Finder tool (http://www.ncbi.nlm.nih.gov/orffinder/) and compared to available nucleotide and amino acid sequences using BLASTn and BLASTx. The molecular weight of the deduced proteins was estimated by the ExPASy Bioinformatics Resource Portal tool (http://web.expasy.org/compute_pi/), and the transmembrane domains were analyzed using TMHMM Server v.2.0 (http://www.cbs.dtu.dk/services/TMHMM/). Repetitive intergenic regions in the MYSV genome that are conserved among rhabdoviruses were analyzed and compared to those in BYSMV. The three component elements of the MYSV intergenic regions (the 3´polyU-rich element, the short non-transcribed variable region, and the 5´conserved element) were identified and aligned using Bioedit software in order to identify the repetitive consensus intergenic region.

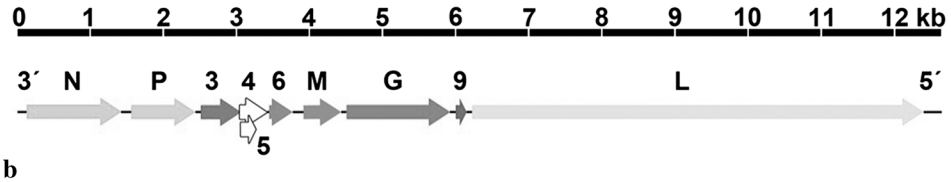
Ten ORFs were predicted in the order 3´ l-N-P-3-4-5-6-M-G-9-L-t 5´, and the MYSV genome organization resembles that of BYSMV and NCMV [9, 10] (Fig. 1a). In particular, MYSV proteins 4 and 5 are predicted to be translated from a single mRNA in an alternate reading frame, a characteristic that is also shared with BYSMV, and these are the only two viruses in the family *Rhabdoviridae* reported to have this feature [9].

The first ORF is 1296 nt long and encodes a putative nucleocapsid protein (N); the second ORF is 879 nt long and encodes a putative phosphoprotein (P protein); the third ORF is 546 nt long and encodes a putative protein 3; the fourth ORF IS 372 nt long and encodes a putative protein 4; the fifth ORF is 219 nt long and encodes protein 5; the sixth ORF is 318 nt long and encodes the putative protein 6; the seventh ORF is 504 nt long and encodes a putative matrix protein (M); the eighth ORF is 1431 nt long and encodes a putative glycoprotein (G); the ninth ORF is 156 nt long and encodes protein 9; and in the last position, the polymerase protein (L) is encoded by ORF 10, which is 6171 nt long (Fig. 1a).

The nucleotide and deduced amino acid sequences of individual ORFs from MYSV genomes of the two isolates were compared using BLASTp analysis. The

Fig. 1 **a.** Schematic representation of the genomic organization of MYSV. Arrows represent ORFs, and bars represent intergenic regions. **b.** Repetitive conserved intergenic regions of the genomic RNA of MYSV compared to BYSMV. *, AACCUUUUCUAUUUUUUGA. The highlighted locations show differences between the viruses. Characteristics of the proteins encoded by the MYSV genome. TM, transmembrane domain predicted by CBS Prediction Servers; ND, not detected; BYSMV, barley yellow striate mosaic virus (GenBank NC028244)

a



b

Location		MYSV Poly (U) rich element	MYSV Variable region	MYSV Conserved element	BYSMV Poly (U) rich element	BYSMV Variable region	BYSMV Conserved element	
"Leader"-N	3´	AUUUUUU	GUGA	CUC	AUUAUUUUU	N19*	CUC	5´
N-P	3´	AUUAUUUUU	GA	CUC	AUUAUUUUU	GA	CUC	5´
P-3	3´	AUUAUUUUU	GA	CUC	AUUAUUUUU	GA	CUC	5´
3-4/5	3´	AUUAUUUUU	GUC	CUC	AUUAUUUUU	GUC	CUC	5´
4/5-6	3´	AUUAUUUUU	GA	CUC	AUUAUUUUU	GA	CUC	5´
6-M	3´	AUUUAUUUU	GA	CUC	AUUUAUUUU	GA	CUC	5´
M-G	3´	AUUUCUUUU	GA	CUC	AUUUAUUUU	GA	CUC	5´
G-9	3´	AUUUCUUUU	GA	CUC	AUUUGUUUU	GA	CUC	5´
9-L	3´	AUUUAUUCAUUUU	GA	CUC	AUUAUUUUU	GA	CUC	5´
L-"trailer"	3´	AUUUAUUUU	GACCU	UAA	AUUAUUUUU	GG	UAA	5´
Consensus	3´	AUUAUUUUU	GA	CUC	AUUAUUUUU	GA	CUC	5´

c

ORF	Calculated mass (kDa)	TM domain	Identities with BYSMV	Query cover	E value	Putative function
1/N	48.3	ND	68%	98%	0.0	Nucleocapsid (N)
2/P	33.4	ND	57%	99%	3e-113	Phosphoprotein (P)
3	20	ND	60%	94%	3e-74	Unknown
4	13.84	ND	37%	95%	9e-19	Unknown
5	8.19	2	44%	97%	7e-09	Unknown
6	11.5	ND	46%	77%	7e-21	Unknown
7/M	19.11	ND	57%	98%	9e-63	Matrix protein (M)
8/G	54.25	ND	57%	97%	0.0	Glycoprotein (G)
9	6.38	1	69%	100%	2e-19	Unknown
10/L	234.7	ND	80%	100%	0.0	Polymerase (L)

maximum nt sequence identity was 99.8% between the ORFs 7 of the two MYSV isolates, and the minimum was 99.2% between the ORFs 3. These results strongly support the hypothesis that the MYSV genomes obtained from maize and wheat correspond to two distinct isolates belonging to the same virus species. Higher heterogeneity at the nucleotide and amino acid level has been reported for isolates of other species of plant rhabdoviruses, using the N gene [11, 12] and the L gene [12, 13].

As shown in Fig 1b, the MYSV genes are separated by conserved repetitive gene junctions, which are characteristic of members of the family *Rhabdoviridae*, flanked by 3´ leader and 5´ trailer sequences. The consensus sequence determined for all intergenic repetitive sequences of both MYSV isolates was the same as that reported for BYSMV [9], although the regions between some pairs of genes differed considerably between MYSV and BYSMV. The intergenic regions showing differences were between the 3' leader and the N gene, M and G, G and 9, 9 and L, and L

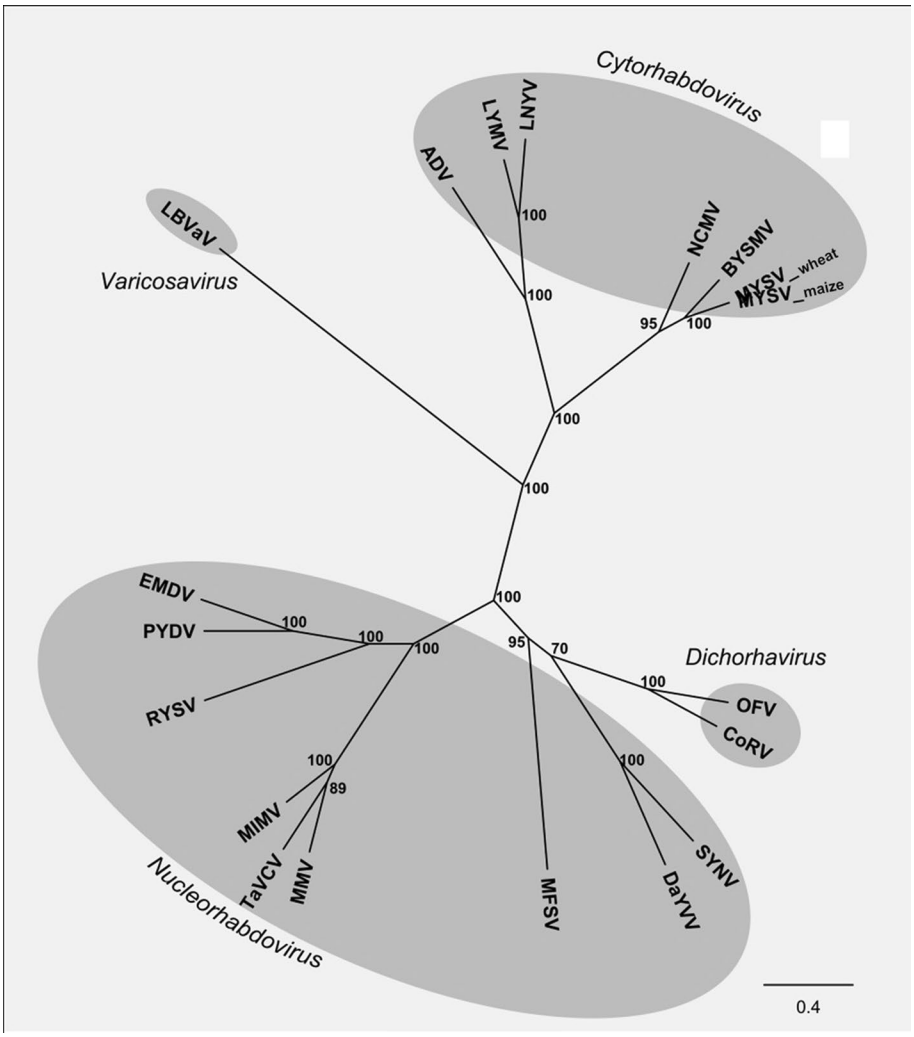


Fig. 2 Phylogenetic analysis of two isolates of MYSV and other plant rhabdoviruses (BYSMV, NC028244; NCMV, NC002251; LNYV, NC007642; LYMV, NC011532; ADV, NC028237; LBVaV, NC011558; EMDV, NC025389; PYDV, NC016136; RYSV, NC003746; MIMV, NC011542; TaVCV, NC006942; MMV, NC005975; MFSV, NC005974; DaYVV, NC028231; SYNV, NC001615; CoRV, KF812526; OFV, NC009609) based on nucleotide sequences of the complete polymerase L gene. The phylogenetic tree was constructed using the maximum-likelihood method (PhyML 3.0)

and the 5´ trailer (Fig. 1b). The 3´ leader and 5´ trailer terminal regions showed sequence complementarity in 15 of the 19 terminal nucleotides in the maize and wheat isolates, giving rise to the typical rhabdovirus panhandle structure [9].

Transmembrane domains were detected in only two of the 10 deduced proteins of the MYSV genome, proteins 5 and 9 (Fig. 1c). This differs from BYSMV, whose protein 8 also contains a transmembrane domain [9]. Using BLASTp analysis, the deduced amino acid sequences of each ORF was evaluated for sequence similarity to annotated proteins of the NCBI non-redundant database. ORFs 1 to 10 were compared with their BYSMV counterparts. The highest amino acid sequence identity between MYSV and BYSMV proteins was 80% between L proteins, and the lowest was 37% between proteins 4 (Fig. 1c).

A phylogenetic analysis was performed based on the complete nucleotide and amino acid (data not shown) sequences of the plant rhabdovirus polymerase gene, using ClustalW software for sequence alignment and PhyML v 3.0 software [14] for tree construction by the maximum-likelihood method with 1000 bootstrap replicates. As expected, the two MYSV isolates could not be differentiated based on the L sequences, a highly conserved gene that is usually used to define species within the family *Rhabdoviridae*, and grouped together in the same branch (Fig. 2). MYSV clustered with BYSMV with a high support (100) in the genus *Cytorhabdovirus*, and the previously reported tree topology for the plant rhabdoviruses group was observed [15–17]. A similar phylogenetic study was carried out using the N gene sequence, with similar results (data not shown).

Based on these results, we consider that the differences observed largely justify the proposal of MYSV being a member of a distinct virus species within the genus *Cytorhabdovirus*, family *Rhabdoviridae*. Yellow striate, affecting maize and wheat crops in Argentina, is an emergent disease that presents a potential economic risk for these widely distributed crops. The availability of the complete genome sequence of MYSV constitutes a useful tool

for epidemiological and biological studies that will contribute to the design of specific diagnostic tools that will aid in defining the real extent of the spread of this disease. In addition, our results will contribute to the development of management strategies and to the design of genetic-engineering-based strategies of control.

Compliance with ethical standards

Funding This work was funded by INTA Research Projects (PNPV-1135022 INTA) and Fundación ArgenINTA.

Conflict of interest Author María Fernanda Maurino has received research grants from Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). Author Analía Delina Dumón has received research grants from Fondo para la Investigación Científica y Tecnológica (FONCyT). Authors Vanina Alemandri, Gabriela Llauger, Luis A. de Haro, M. Fernanda Mattio, Mariana del Vas, Irma Graciela Laguna and María de la Paz Giménez Pecci declare that they have no conflict of interest.

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

References

1. USDA (2017) United States Department of Agriculture. In: Prod. Supply Distrib. Online. https://apps.fas.usda.gov/psdonline/app/index.html#/app/downloads
2. Maurino MF, Laguna G, Giolitti F et al (2012) First ocurrence of a rhabdovirus infecting maize in Argentina. Plant Dis 96:1383
3. Dumón AD, Mattio MF, Argüello-Caro EB et al (2015) Occurrence of a closely-related isolate to Maize yellow striate virus in wheat plants. Agriscientia 32:107–112
4. Tritt A, Eisen JA, Facciotti MT, Darling AE (2012) An integrated Pipeline for de novo assembly of microbial genomes. PLoS One. doi:10.1371/journal.pone.0042304
5. Kearse M, Moir R, Wilson A et al (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649. doi:10.1093/bioinformatics/bts199
6. Milne I, Stephen G, Bayer M et al (2013) Using Tablet for visual exploration of second-generation sequencing data. Brief Bioinform 14:193–202. doi:10.1093/bib/bbs012
7. Lander ES, Waterman MS (1988) Genomic mapping by fingerprinting random clones: a mathematical analysis. Genomics 2:231–239
8. Frohman MA, Dush MK, Martin GR (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc Natl Acad Sci USA 85:8998–9002. doi:10.1073/pnas.85.23.8998
9. Yan T, Zhu J-R, Di D et al (2015) Characterization of the complete genome of *Barley yellow striate mosaic virus* reveals a nested gene encoding a small hydrophobic protein. Virology 478:112–122
10. Tanno F, Nakatsu A, Toriyama SS, Koyima M (2000) Complete nucleotide sequence of Northern cereal mosaic virus and its genome organization. Arch Virol 145:1373–1384
11. Callaghan B, Dietzgen RG (2005) Nucleocapsid gene variability reveals two subgroups of *Lettuce necrotic yellows virus*. Arch Virol 150:1661–1667. doi:10.1007/s00705-005-0528-7
12. Revill P, Trinh X, Dale J, Harding R (2005) *Taro vein chlorosis virus*: characterization and variability of a new *Nucleorhabdovirus*. J Gen Virol 86:491–499. doi:10.1099/vir.0.80591-0
13. Klerks MM, Lindner JL, Vaskova D et al (2004) Detection and tentative grouping of *Strawberry crinkle virus* isolates. Eur J Plant Pathol 110:45–52
14. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52:696–704. doi:10.1080/10635150390235520
15. Dietzgen RG, Kuhn JH, Clawson AN et al (2014) Dichorhavirus: a proposed new genus for Brevipalpus mite-transmitted, nuclear, bacilliform, bipartite, negative-strand RNA plant viruses. Arch Virol 159:607–619. doi:10.1007/s00705-013-1834-0
16. Kondo H, Maeda T, Shirako Y, Tamada T (2006) Orchid fleck virus is a rhabdovirus with an unusual bipartite genome. J Gen Virol 87:2413–2421. doi:10.1099/vir.0.81811-0
17. Ramalho TO, Figueira AR, Sotero AJ et al (2014) Characterization of Coffee ringspot virus-Lavras: a model for an emerging threat to coffee production and quality. Virology 464–465:385–396. doi:10.1016/j.virol.2014.07.031