Arch Virol (2017) 162:2481–2484
DOI 10.1007/s00705-017-3363-8

ANNOTATED SEQUENCE RECORD

Molecular characterization of yerba mate chlorosis-associated virus, a putative cytorhabdovirus infecting yerba mate (*Ilex paraguariensis*)

Nicolás Bejerman[1,3] · Soledad de Breuil[1,3] · Humberto Debat[1] · Marcos Miretti[2,3] · Alejandra Badaracco[3] · Claudia Nome[1]

Received: 6 January 2017 / Accepted: 2 April 2017 / Published online: 27 April 2017

Abstract We present the molecular characterization of a new virus infecting yerba mate (*Ilex paraguariensis* St. Hil.) in Argentina. Deep sequencing of diseased yerba mate plants showing chlorotic linear patterns, chlorotic rings, and vein yellowing resulted in the identification of a new virus resembling plant rhabdoviruses in sequence and genome structure. We have determined the complete genome sequence of this virus, which is 12,876 nt long. Seven open reading frames (ORFs) were identified in the antigenomic orientation of the negative-sense, single-stranded viral RNA, in the order 3′-N-P-P3-P4-M-G-L-5′. Phylogenetic analysis suggested that the described virus is a new member of the genus *Cytorhabdovirus*, which was supported by the observation of rhabdovirus-like particles within the cytoplasm of infected yerba mate cells. The virus has been tentatively named "yerba mate chlorosis-associated virus" (YmCaV). The availability of the YmCaV genome sequence will contribute to assessing the genetic variability of this virus and determining its role in this yerba mate disease.

Yerba mate (*Ilex paraguariensis* St. Hil., Aquifoleaceae) is one of the most important subtropical trees/shrubs in Argentina, Brazil, and Paraguay. Its leaves and stems are widely utilized in the preparation of an infusion popularly known as "mate". In Argentina, the area cultivated with this tree is about 165,200 ha with a total yield of 806,324 t [1]. In 2010, yerba mate plants showing chlorotic linear patterns, chlorotic rings, and vein yellowing were observed in the northeast region of Argentina (Supp. Fig. 1A).

In this report, we describe the molecular characterization of a rhabdovirus associated with these symptoms in Argentina, tentatively named "yerba mate chlorosis-associated virus" (YmCaV). This virus can be transmitted to healthy yerba mate plants by grafting; however, it cannot be transmitted by mechanical inoculation.

Yerba mate plants showing the described symptoms were collected near Cerro Azul, Misiones Province, Argentina, and kept in a glasshouse at IPAVE-CIAP, INTA. Total RNA was extracted using a Quick-RNA™ MiniPrep kit (Zymo Research, Irvine, USA) according to the manufacturer's instructions. Total RNA was sent to Fasteris Life Sciences SA (Plan-les-Ouates, Switzerland) for small-RNA sequencing, and bands located between 21 and 30 nt were excised, purified, processed and sequenced on an Illumina HiSeq 2000 Genome Analyzer. The raw data were processed to remove barcode adaptors, and low-quality nucleotide sequences were trimmed using NGS CRUMBS (http://bioinf.comav.upv.es/ngs_crumbs) [2]. The cleaned reads were assembled *de novo* using the software package Velvet v0.6.04 [3], and individual contigs were analyzed using BLASTx. Five assembled

Electronic supplementary material The online version of this article (doi:10.1007/s00705-017-3363-8) contains supplementary material, which is available to authorized users.

✉ Nicolás Bejerman
nicobejerman@gmail.com

[1] Instituto de Patología Vegetal, Centro de Investigaciones Agropecuarias, Instituto Nacional de Tecnología Agropecuaria (IPAVE-CIAP-INTA), Camino 60 Cuadras Km 5,5, X5020ICA Córdoba, Argentina

[2] Grupo de Investigación en Genética Aplicada, Instituto de Biología Subtropical (GIGA, IBS, FCEQyN, UNaM - CONICET) Jujuy 1745, N3300NFK Posadas, Misiones, Argentina

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

sequence contigs of 1,318 bp, 833 bp, 375 bp, 4,675 and 3,318 bp were obtained (Fig. 1A). Next, reverse transcription PCR (RT-PCR), cloning, and Sanger sequencing using specific primers designed from the assembled sequence (Fig. 1B) were employed to close the gaps between the five contigs, whereas the ends were obtained using the 3′ and 5′ RACE System for Rapid Amplification of cDNA Ends (Life Technologies). This resulted in a sequence of 12,876 nucleotides (nt) covering the whole genome of YmCaV (GenBank accession number KY366322) (Fig. 1C). Sequences were compiled and analyzed using the Lasergene 10 software package (DNASTAR Inc., Madison, WI, USA). The percentage amino acid (aa) sequence identity of the predicted ORFs of YmCaV to available cyto- and nucleorhabdovirus genome sequences was calculated using ClustalW [4] implemented in the BioEdit 7.0.9.0 software [5]. Phylogenetic analysis of the nt sequence of the complete L protein coding region was carried out using the maximum-likelihood (ML) method, as implemented in MEGA 6.0 [6], under the best-fit substitution model. Protein sequences were scrutinized for transmembrane domains, using the TMHMM version 2.0 tool (http://www.cbs.dtu.dk/services/TMHMM/) [7].

The coding region of YmCaV is 12,379 nt in length and contains seven open reading frames (ORFs) in the anti-genomic strand (Fig. 1C). BlastX analysis of the identified ORFs showed that they code for a nucleocapsid protein (N; ORF1), phosphoprotein (P; ORF2), movement protein (P3; ORF3), matrix protein (M; ORF 5), glycoprotein (G; ORF 6), and RNA-dependent RNA polymerase (L; ORF 7) based on high sequence identity scores with plant rhabdoviruses, whereas P4 (ORF4) did not have significant matches with any plant rhabdovirus entries in GenBank (Supplemental Table 1). The molecular characterization of proteins encoded by YmCaV genome as determined by predictive algorithms are shown in Supplemental Table 1. The coding sequences are flanked by complementary 3' leader (l) and 5' trailer (t) sequences, which are 151 and 346 nt long, respectively, revealing a genome organization of 3' l–N–P–P3–P4-M–G–L–t 5' (Fig. 1C). The predicted genome structure resembles the basic canonical organization described for plant rhabdoviruses [8, 9]; however, an additional ORF, named as P4, was found in the YmCaV genome. Accessory genes located at various positions in the genome have commonly been described for several other animal and plant rhabdoviruses [10]. Nevertheless, YmCaV is only the third plant rhabdovirus that has been found to contain an accessory gene located between the P3 and M genes. The cytorhabdoviruses barley yellow striate mosaic virus (BYSMV) and northern cereal mosaic virus (NCMV) have two or three accessory genes in this position, and their function remains unknown [10–12]. In addition, a nested gene named P5 was found to be encoded by BYSMV in an alternative frame within gene 4, expressed via a leaky scanning mechanism [12]. YmCaV P4 has 67 amino acid residues with a predicted molecular weight of 7.8 kDa and an isoelectric point of 9.22 (Supplemental Table 1). A transmembrane domain was identified in the P4 N-terminal region (aa positions 11-33) (Supplemental Table 1), which may indicate that P4 is a membrane-associated protein. YmCaV P4 is a small hydrophobic (SH) protein that is reminiscent of the BYSMV SH P5. On the other hand, NCMV P4, P5 and P6 are basic proteins like YmCaV P4; however, no transmembrane domain was found in any these proteins.

Four transmembrane domains were identified in the YmCaV putative glycoprotein G (aa positions 5-27, 37-59, 80-97, 560-582) (Supplemental Table 1). Similar domains have been reported in the G proteins of other plant rhabdoviruses [12–16], which is consistent with their membrane functions.

Like all plant rhabdoviruses, YmCaV genes are separated by intergenic "gene junction" regions, which are

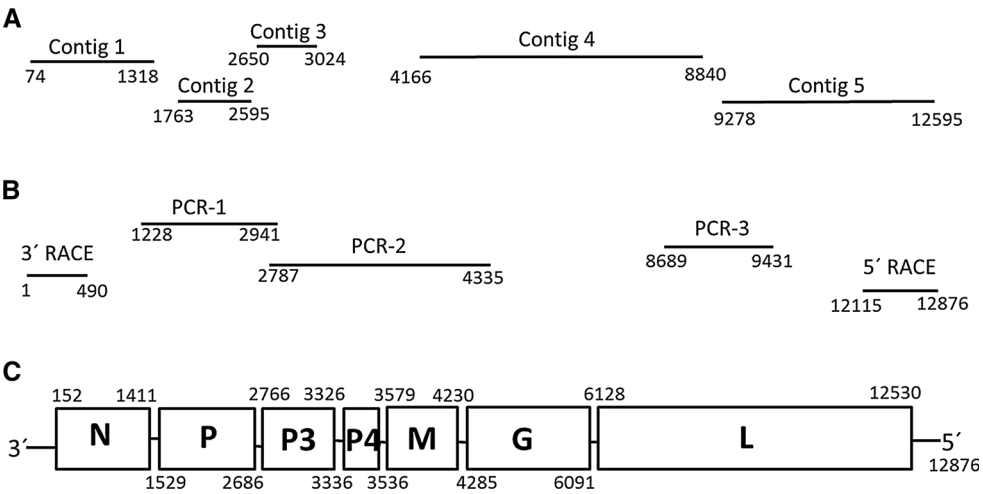

Fig. 1 Diagrammatic representation of the contigs from *de novo* assembly of NGS data (A), products of RT-PCR (B), and the putative genome organization of yerba mate chlorosis-associated virus (YmCaV) (C). Positions are marked at extremities

composed of the polyadenylation signal of the preceding gene, a short intergenic region, and the transcriptional start of the following gene (Supplemental Table 2). The YmCaV consensus "gene junction" region sequence is 3′ AUUAUUUUUUGACUC 5′ and is almost identical to that of BYSMV (Supplemental Table 2). Nevertheless, unlike BYSMV and NCMV, no "gene junction" region could be identified between the P3 and P4 genes of YmCaV. The significance of this finding is unknown.

Amino acid sequence comparisons between the deduced YmCaV proteins and the corresponding sequences of other plant rhabdoviruses (Supplemental Table 3) revealed a close relationship to cytorhabdoviruses, in particular with BYSMV and NCMV. However, the aa sequence identity between YmCaV and BYSMV was low and ranged from 12.4% in the P protein to 32.3% in the L polymerase protein. Moreover, the YmCaV P protein was more similar to the P protein of lettuce necrotic yellows virus (LNYV) (Supplemental Table 3). A low level of sequence identity is common between individual plant rhabdoviruses; these viruses display a high level of diversity in both their sequence and genome organization [9, 10]. Therefore, the sequence of this new plant rhabdovirus will provide new information about their diversity and the variability of their genomic organization, leading toward a better understanding of their evolution.

A phylogenetic tree based on the complete nt sequence of the polymerase gene ORF shows that YmCaV clusters together with other viruses in the genus *Cytorhabdovirus* (Fig. 2), with a close evolutionary relationship with BYSMV and NCMV, which are vectored by planthoppers, and colocasia bobone disease-associated virus (CBDaV) (Fig. 2), which is also likely to be vectored by planthoppers [17]. The vector of YmCaV has not been identified yet, but based on the phylogenetic analysis, it is tempting to speculate that a planthopper may be the vector of this virus. The phylogenetic clustering agrees with the sequence identities detailed in Supplemental Table 3, showing that BYSMV, NCMV and CBDaV are the cytorhabdoviruses that are most closely related to YmCaV. Like BYSMV and NCMV, the YmCaV genome has an additional ORF between its P3 and M genes; however, it has only one additional ORF, whereas BYSMV and NCMV, which infect monocots, have three additional ORFs in this location [12, 13].

Sequence comparisons and phylogenetic analysis indicated that YmCaV should be considered a representative of a putative new species in the genus *Cytorhabdovirus,* family *Rhabdoviridae*. Ultrathin sections of symptomatic yerba leaves were observed in a transmission electron microscope. Rhabdovirus-like particles were always found within the cytoplasm of mesophyll cells at very low concentrations (Supplemental Fig. 1B), which further supports the classification of YmCaV as a tentative cytorhabdovirus.

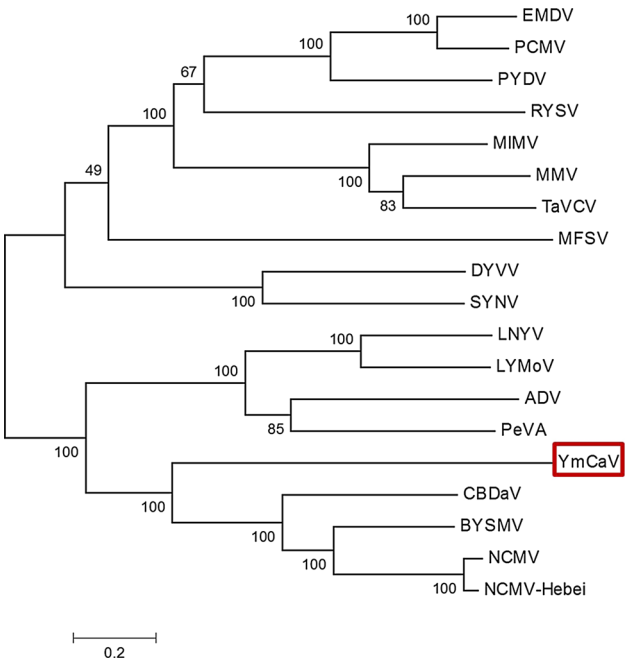

Fig. 2 Phylogenetic tree based on the alignment of the L polymerase open reading frame nucleotide sequences of YmCaV and other plant rhabdoviruses. The tree was constructed in MEGA 6 using the maximum-likelihood (ML) method with the GTR + G + I model. The values on the branches show the percentage of support out of 1000 bootstrap replications. The tree is rooted at the midpoint; nucleorhabdovirus and cytorhabdovirus clades are indicated by brackets. The scale bar indicates the number of substitutions per site. The viruses used to construct the tree, and their accession numbers are as follows: alfalfa dwarf virus (ADV; KP205452), barley yellow striate mosaic virus (BYSMV; KM213865), colocasia bobone disease associated-virus (CBDaV; KT381973), datura yellow vein virus (DYVV; KM823531), eggplant mottled dwarf virus (EDMV; NC_025389), lettuce yellow mottle virus (LYMoV; EF687738), lettuce necrotic yellows virus (LNYV; NC_007642); maize fine streak virus (MFSV; AY618417), maize Iranian mosaic virus (MIMV; DQ186554), maize mosaic virus (MMV; AY618418), northern cereal mosaic virus (NCMV; AB030277 and GU985153), persimmon virus A (PeVA; NC_018381), physostegia chlorotic mottle virus (PCMV; KX636164), potato yellow dwarf virus (PYDV; GU734660), rice yellow stunt virus (RYSV; NC_003746); sonchus yellow net virus (SYNV; L32603), taro vein chlorosis virus (TaVCV; AY674964)

The availability of the YmCaV genome sequence will now contribute to assessing the genetic variability of this virus and to studying its role in yerba mate disease.

Acknowledgements We thank Dr. Ralf Dietzgen (University of Queensland) for a critical reading of the manuscript. This work was supported by INTA project PNIND PE 1108072 and FONCYT grant (PICT 2014-1212).

Compliance with ethical standards

Conflict of interest There is no conflict of interest.

References

1. https://datos.magyp.gob.ar. Accessed 30th November 2016
2. http://bioinf.comav.upv.es/ngs_crumbs. Accessed 15th August 2016
3. Zerbino DR, Birney E (2008) Velvet: algorithms for de novo read assembly using de Bruijn graphs. Genome Res 18:821–829
4. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680
5. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/ NT. Nucleic Acids Symp Ser 41:95–98
6. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729
7. Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567–580
8. Dietzgen RG, Calisher CH, Kurath G, Kuzmin IV, Rodriguez LL, Stone DM, Tesh RB, Tordo N, Walker PJ, Wetzel T, Whitfield AE (2011) Family *Rhabdoviridae*. In: King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (eds) Virus taxonomy, 9th edn. Elsevier, San Diego, pp 686–714
9. Mann K, Dietzgen RG (2014) Plant rhabdoviruses: new insights and research needs in the interplay of negative-strand RNA viruses with plant and insect hosts. Arch Virol 159:1889–1900
10. Walker PJ, Dietzgen RG, Joubert DA, Blasdell KR (2011) Rhabdovirus accessory genes. Virus Res 162:110–125
11. Tanno F, Nakatsu A, Toriyama S, Kojima M (2000) Complete nucleotide sequence of Northern cereal mosaic virus and its genome organization. Arch Virol 145:1373–1384
12. Yan T, Zhu JR, Di D, Gao Q, Zhang Y, Zhang A, Yan C, Miao H, Wang XB (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
13. Bandyopadhyay A, Kopperud K, Anderson G, Martin K, Goodin M (2010) An integrated protein localization and interaction map for Potato yellow dwarf virus, type species of the genus Nucleorhabdovirus. Virology 402:61–71
14. Bejerman N, Giolitti F, de Breuil S, Trucco V, Nome C, Lenardon S, Dietzgen RG (2015) Complete genome sequence and integrated protein localization and interaction map for alfalfa dwarf virus, which combines properties of both cytoplasmic and nuclear plant rhabdoviruses. Virology 483:275–283
15. Dietzgen RG, Callaghan B, Wetzel T, Dale JL (2006) Completion of the genome sequence of Lettuce necrotic yellows virus, type species of the genus *Cytorhabdovirus*. Virus Res 118:16–22
16. Tsai CW, Redinbaugh MG, Willie KJ, Reed S, Goodin M, Hogenhout SA (2005) Complete genome sequence and in planta subcellular localization of maize fine streak virus proteins. J Virol 79:5304–5314
17. Higgins CM, Bejerman N, Li M, James A, Dietzgen RG, Pearson MN, Revill PA, Harding RM (2016) Complete genome sequence of Colocasia bobone disease-associated virus, a putative cytorhabdovirus infecting taro. Arch Virol 161:7545–7548