Archives of Virology (2018) 163:2811–2815
https://doi.org/10.1007/s00705-018-3910-y

BRIEF REPORT

Identification and molecular characterization of a novel circular single-stranded DNA virus associated with yerba mate in Argentina

Nicolás Bejerman[1,2] · Soledad de Breuil[1,2] · Claudia Nome[1]

Received: 6 March 2018 / Accepted: 27 April 2018 / Published online: 6 June 2018

Abstract

A single-stranded DNA (ssDNA) virus was detected in Yerba mate samples showing chlorotic linear patterns, chlorotic rings and vein yellowing. The full-genome sequences of six different isolates of this ssDNA circular virus were obtained, which share > 99% sequence identity with each other. The newly identified virus has been tentatively named as yerba mate-associated circular DNA virus (YMaCV). The 2707 nt-long viral genome has two and three open reading frame on its complementary and virion-sense strands, respectively. The coat protein is more similar to that of mastreviruses (44% identity), whereas the replication-associated protein of YMaCV is more similar (49% identity) to that encoded by a recently described, unclassified ssDNA virus isolated on trees in Brazil. This is the first report of a circular DNA virus associated with yerba mate. Its unique genome organization and phylogenetic relationships indicates that YMaCV represents a distinct evolutionary lineage within the ssDNA viruses and therefore this virus should be classified as a member of a new species within an unassigned genus or family.

Yerba mate (*Ilex paraguariensis* St. Hil., Aquifoliaceae) is one of the most important subtropical trees or shrub crops in Argentina, Brazil, and Paraguay. Its leaves and stems are widely used in an infusion popularly known as "mate". In Argentina, the cultivated area with this tree is about 165,200 ha with a total yield of 775 t per year [1]. In 2010, yerba mate plants showing chlorotic linear patterns, chlorotic rings, and veins yellowing in symptomatic leaves, were observed in the northeast region of Argentina [2]. Deep sequencing of small RNAs of symptomatic yerba mate plants collected in Cerro Azul, Misiones Province, Argentina, revealed the presence of a new virus: cytorhabdovirus yerba mate chlorosis associated virus (YmCaV) [2]. Two assembled sequences (contigs) showed a low degree of similarity to the capsid and replication associated proteins encoded by mastreviruses and the ssDNA virus temperate fruit decay-associated virus (TFDaV), respectively, based on BlastX analysis.

In this study we describe the complete genome sequence of a new circular single strand DNA (ssDNA) virus isolated from yerba mate in Argentina, which was named yerba mate-associated circular DNA virus (YMaCV).

Leaf samples with the described symptoms were collected in six different locations around Corrientes and Misiones Provinces in Argentina (Supplementary Figure 1). DNA extractions were performed using 100 mg of infected tissue with the cetyltrimethylammonium bromide (CTAB) method [3]. Total DNA from each isolate was subjected to Phi29 DNA polymerase treatment (TempliPhi™, GE Healthcare, USA) to obtain a copy of the complete genome of the ssDNA virus by rolling circle amplification (RCA). A linear DNA fragment of about 2.7 kb was generated with *Bam*HI digestion, cloned and three clones per isolate were sequenced bi-directionally with the universal M13 primers. Additionally, primers were designed, by primer-walking, to sequence the full genome of each clone. The sequence fragments obtained for each plasmid DNA were edited and

Handling Editor: Elvira Fiallo-Olivé.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00705-018-3910-y) 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] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Godoy Cruz 2290 (1425 FBQ) CABA, Buenos Aires, Argentina

assembled to obtain the complete genome sequence of each isolate of the ssDNA virus, which were deposited in GenBank under accession numbers MG748715-748720 (Supplementary Figure 1), using Geneious R9.1 (Boimatters, New Zealand).

The genome of each YMaCV isolate was determined to be 2,707 nucleotides (nt) in size and its circular DNA contains a nonanucleotide sequence CATTATTAC that indicates the probable origin of virion-strand and DNA replication. This sequence differs in two nucleotides from the nonanucleotide sequence of TFDaV, nanoviruses, alphasatellites and circoviruses (TAGTATTAC) [4]. Five open reading frames (ORFs) that might encode putative proteins were identified. The inferred genome organization has three virion-sense (*V1*, *V2* and *V3*) and two complementary-sense (*C1* and *C2*) genes with two intergenic regions (Fig. 1). Similar to curto-and becurtoviruses [5], YMaCV has three ORFs in the virion-sense strand with the same genomic location. However, the number and genomic location of the ORFs in the complementary-sense differs in YMaCV, thus the genomic organization of this virus is distinct from those of all other known ssDNA viruses. The genes on sense and complementary strands are separated by two intergenic regions (Fig. 1) that have similar size.

A BLASTp (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome) analysis of the predicted proteins of YMaCV showed that C2, V2 and V3 proteins had no significant sequence similarity to any others sequences in the database. However, similar to the putative movement protein of TFDaV, as well as several geminiviruses and nanoviruses where transmembrane domains have been predicted [4, 6], two transmembrane domains were identified in the V2 protein sequence. Furthermore, the ORF V2 (342 nt; 113aa; 13.32 kDa) is immediately upstream of ORF V1, putatively coding for the capsid protein (CP), a location where movement proteins are usually encoded within the monopartite members of the *Geminiviridae* family [5]. The ORF V1 which is 792 nt long, and encodes a 29.73 kDA protein of 263aa is most similar to the CP encoded by the mastrevirus chickpea chlorosis virus-A (CpCV-A) (44% identity; 93% query coverage; E value, 7e-56). Unlike the V1 protein, the translation product of ORF C1 (960 nt; 319 aa; 36.61 kDa) is most closely related to the replication associated protein (Rep) encoded by TFDaV (49% identity; 95% query coverage; E value, 9e-94), which is an unclassified ssDNA virus discovered recently on fruit trees in Brazil [4]. YMaCV C1 protein is also homologous to Rep proteins found in a wide variety of ssDNA circular replicons including nanoviruses, alphasatellites and some currently unclassified environmental ssDNAs. Therefore V1, C1 and V2 proteins were identified as the putative CP, Rep and MP proteins.

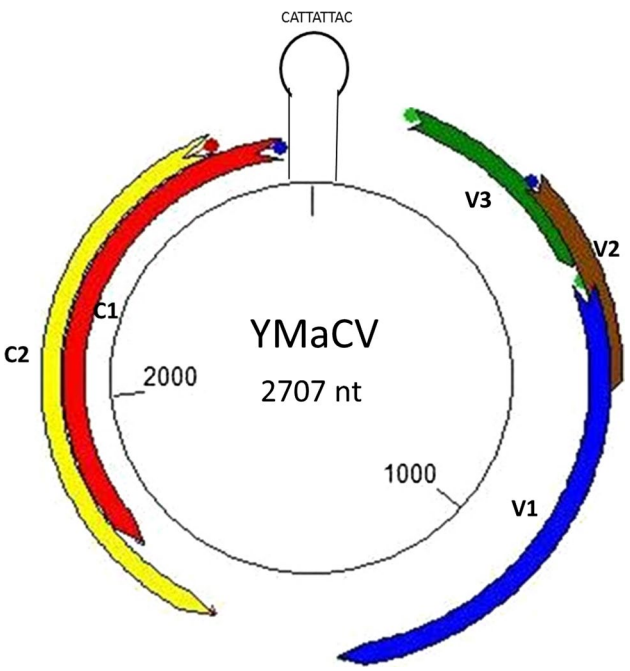

Fig. 1 Genomic organization of the yerba mate-associated circular virus (YMaCV) as well as the stem-loop structure with the nonanucleotide sequence containing the probable virion-strand origin of replication

Genome-wide pairwise comparisons using Sequence Demarcation Tool (SDT v1.2) [7] indicated that the six isolates of YMaCV share 99.0 to 99.9% identity with each other, which likely means that this virus emerged recently. Moreover these genome-wide pairwise comparisons showed that YMaCV shared the highest identity (59.8%) with the unclassified ssDNA virus TFDaV. In pairwise comparisons, the YMaCV CP aa sequence shared highest identity (44.2%) with the mastrevirus CpCV-A (Fig. 2A); whereas YMaCV Rep protein sequence was most similar (49.4%) to the corresponding sequence of the unclassified ssDNA virus TFDaV (Fig. 2B). These results showed that there is very low variability within the YMaCV isolates and that this virus is distantly related to other ssDNA viruses.The low genetic identity of YMaCV with other ssDNA viruses suggests that this virus should be classified as a member of a new species.

In order to determine the phylogenetic relationships of YMaCV with other ssDNA viruses and to unravel its possible evolutionary history, maximum likelihood (ML) phylogenetic analyses were carried out using the aa sequences of the CP and Rep proteins. The predicted YMaCV CP and Rep aa sequences were aligned with the corresponding sequences of ssDNA viruses described in Supplementary Table 1, using MUSCLE [8] as implemented in Mega 7 [9]. ML phylogenetic trees were inferred using Mega 7 [9] with LG +G+I and RtRev+G+F+I amino acid substitution models, respectively, chosen as the best-fit using ProtTest [10]

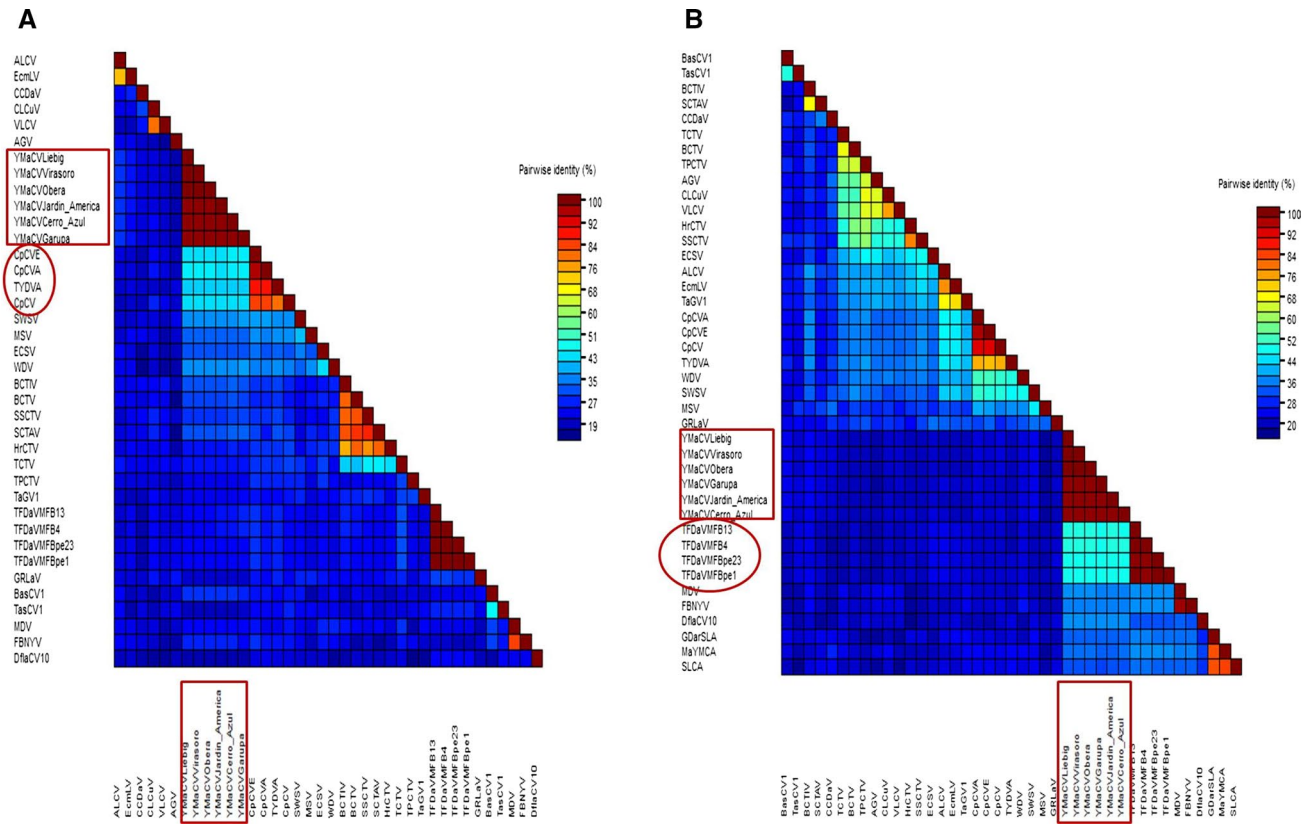

Fig. 2 Pairwise identity matrix of yerba mate-associated circular virus (YMaCV) and representative ssDNA virus CP (A) and Rep (B) amino acid sequences, generated using the SDT v1.2 software (Muhire et al., 2014). Acronyms and accession numbers of every virus are listed in Supplementary Table 1

together with 1000 bootstrap replicates. We used the Akaike Information Criterion (AIC) to select the best-fitting evolution model. The phylogenetic analysis performed with the CP aa sequence showed that the YMaCV isolates are located in a monophyletic clade which clusters with the CP of the mastreviruses CpCV, CpCV-A, CpCV-E as well as tobacco yellow dwarf virus-A (TYDV-A) (Fig. 3A); whereas the Rep protein based tree showed that YMaCV isolates formed a distinct clade that clustered with the Rep protein of the unclassified ssDNA virus TFDaV (Fig. 3B); this cluster is also related to the Rep proteins of nanoviruses and alphasatellites, which further supports the results obtained when the CP and Rep sequences were analyzed using SDT. Phylogenetic analysis placed YMaCV in a unique taxon within ssDNA viruses; therefore it can be concluded that this virus represents an evolutionary distinct ssDNA virus member. Furthermore, the different evolutionary relationships of the Rep and CP proteins likely support a modular organization and a different evolutionary history of the virion-sense and complementary-sense frame of the YMaCV genome.

It is evident that the YMaCV CP and Rep phylogenies are not congruent, probably due to an inter-family recombination resulting in a chimeric virus encoding Rep and CP with differential evolutionary histories. However, when recombination analysis of full-length genomes was performed using RDP, GENECONV, MaxChi, BootScan, 3Seq, Chimera and SiScan statistical methods, implemented in the RDP4.67 software [11], no recombination events were detected in the YMaCV sequence. This result could be explained by the fact that the non-recombinant descendants of YMaCV parents have not been described or no longer exist.

In this study we identified and characterized a novel ssDNA virus infecting yerba mate in Argentina, which was named as yerba mate-associated circular DNA virus (YMaCV). This virus displays a geminivirus-like genomic size and organization, and encodes a CP that is most closely related to that of mastreviruses, viruses which taxonomically belong to the *Geminiviridae* family; however, the Rep protein is most closely related to unrelated viruses not and the YMaCV nonanucleotide sequences are different from those characteristic of geminiviruses. Thus, this virus represents an evolutionary distinct lineage within the ssDNA viruses and may represent a new genus within a new family of ssDNA viruses that could also include TFDaV.

The use of next generation sequencing (NGS) has identified a large number of whole genome sequences

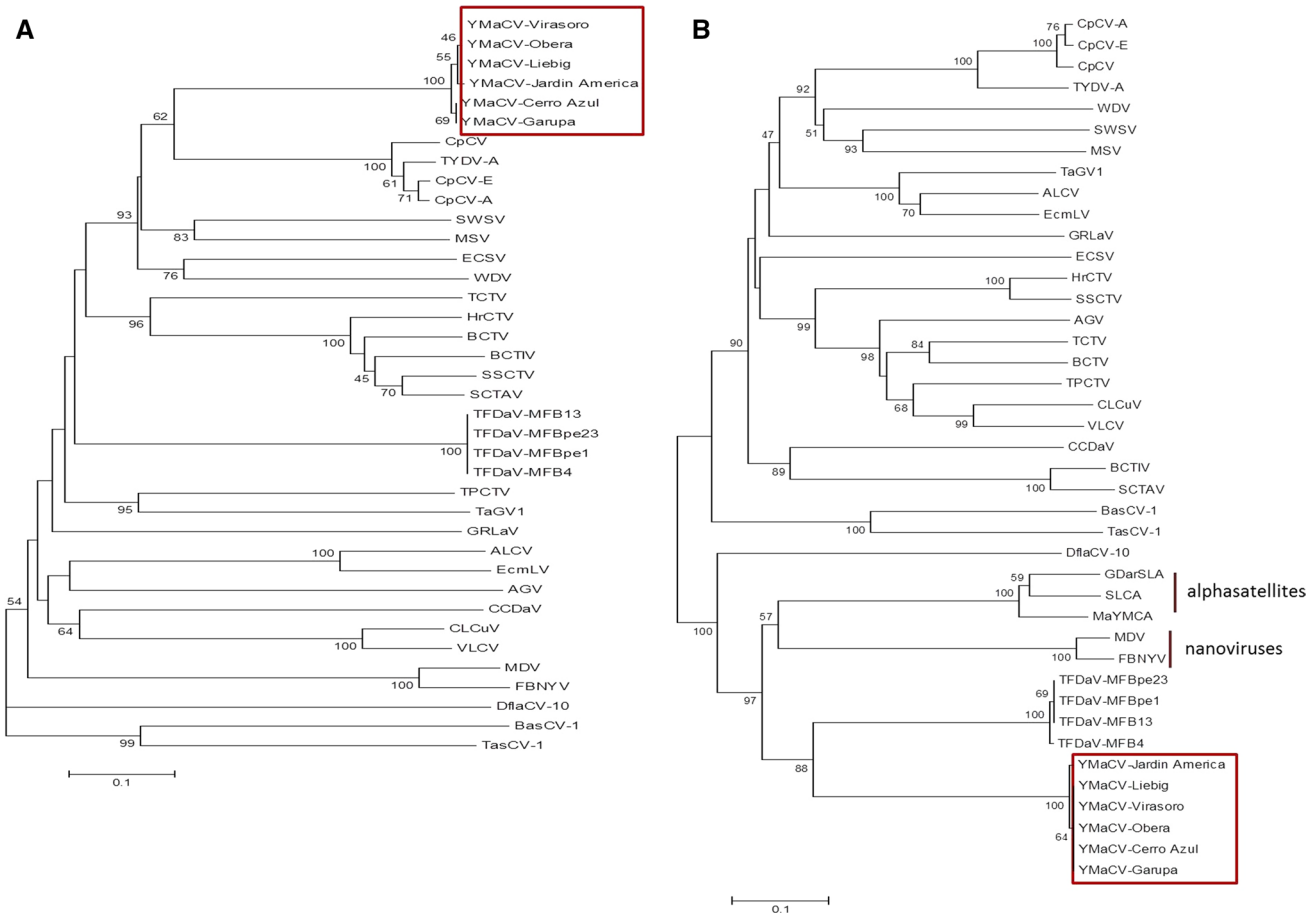

Fig. 3 Maximum likelihood phylogenetic tree analyzing the yerba mate-associated circular virus (YMaCV) and representative ssDNA virus CP (C) and Rep (D) amino acid sequences, generated using Mega7 software with LG+I+G and rtREV+I+G+F models, respectively. Bootstrap values following 1000 replicates are given at the nodes, but only values above 50% are shown. The bar below each tree represents substitutions per site. Acronyms and accession numbers of every virus used to construct these phylogenetic trees are listed in Supplementary Table 1

for viruses with chimeric genome architecture and phylogeny, such as YMaCV. Issues regarding taxonomical classification of ssDNA viruses that are known only from metagenomic data have been recently discussed, where a new robust framework for sequence-based virus taxonomy was identified as crucial for the comprehensive characterization of the global virome [12]. Furthermore, biological properties of most of these viruses have not been characterized, therefore the information in this paper will also be useful to elucidate the taxonomical classification of those currently unassigned ssDNA viruses.

This is the first report of a circular ssDNA virus infecting yerba mate, both in Argentina and worldwide. Further work, such as the construction of an infectious clone, is necessary to unravel not only the symptoms induced by YMaCV, but also its role in yerba mate chlorotic disease, which will be useful to understand the biological significance of this virus in infected plants.

Acknowledgements This work was supported by INTA project PNIND PE 1108072 and FONCYT grant (PICT 2014-1212).

Compliance with ethical standards

Conflict of interest All authors 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. https://datos.magyp.gob.ar. Accessed 15 Jan 2018
2. Bejerman N, de Breuil S, Debat H, Miretti M, Badaracco A, Nome C (2017) Molecular characterization of yerba mate chlorosis-associated virus, a putative cytorhabdovirus infecting yerba mate (*Ilex paraguariensis*). Arch Virol 162:2481–2484

3. Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15
4. Basso MF, da Silva JC, Fajardo TV, Fontes EP, Zerbini FM (2015) A novel, highly divergent ssDNA virus identified in Brazil infecting apple, pear and grapevine. Virus Res 210:27–33
5. Varsani A, Roumagnac P, Fuchs M, Navas-Castillo J, Moriones E, Idris A, Briddon RW, Rivera-Bustamante R, Zerbini FM, Martin DP (2017) Capulavirus and Grablovirus: two new genera in the family Geminiviridae. Arch Virol 162:1819–1831
6. Ma Y, Navarro B, Zhang Z, Lu M, Zhou X, Chi S, Di Serio F, Li S (2015) Identification and molecular characterization of a novel monopartite geminivirus associated with mulberry mosaic dwarf disease. J Gen Virol 96:2421–2434
7. Muhire BM, Varsani A, Martin DP (2014) SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. PLoS One 9:e108277
8. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797
9. Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis Version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874
10. Abascal F, Zardoya R, Posada D (2005) ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21:2104–2105
11. 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
12. Simmonds P, Adams MJ, Benko M, Breitbart M, Brister JR, Carstens EB, Davison AJ, Delwart E, Gorbalenya AE, Harrach B, Hull R, King AMQ, Koonin EV, Krupovic M, Kuhn JH, Lefkowitz EJ, Nibert ML, Orton R, Roossinck MJ, Sabanadzovic S, Sullivan MB, Suttle CA, Tesh RB, van der Vlugt RA, Varsani A, Zerbini FM (2017) Consensus statement: virus taxonomy in the age of metagenomics. Nat Rev Microbiol. https://doi.org/10.1038/nrmicro.2016.1177