

The complete genome of a putative endornavirus identified in yerba mate (*Ilex paraguariensis* St. Hil.)

Humberto J. Debat · Mauro Grabiele · Patricia M. Aguilera ·
Rosana Bubillo · Pedro D. Zapata · Dardo A. Marti ·
Daniel A. Ducasse

Received: 11 April 2014 / Accepted: 7 June 2014
© Springer Science+Business Media New York 2014

Abstract We present the first report of a virus infecting the subtropical tree crop yerba mate (*Ilex paraguariensis* St. Hil.). Total RNA purification, followed by next-generation sequencing, transcripts assembly and annotation, resulted in the identification of a new endornavirus species infecting yerba mate. The complete sequence of the linear dsRNA viral genome is 13,954-nt long, contains a single 13,743 nt ORF, and presents a 149 nt 5'UTR and a 61 nt 3'UTR. The predicted ORF encodes a 4,581 aa polypeptide with a UDP-glucose glycosyl-transferase, a capsular polysaccharide synthesis protein, and a RNA-dependent RNA polymerase domain. The name yerba mate endornavirus is proposed for the identified virus. Due to the intriguing

peculiarities of this virus family, and the complete lack of the yerba mate virus literature, we consider that the information reported here will be helpful in leading to a new and needed attention to this important topic and crop.

Keywords Yerba mate · *Ilex paraguariensis* · Endornavirus · Virus genome · Phylogenetic analysis

Introduction

Yerba mate (*Ilex paraguariensis* St. Hil.) is one of the most important subtropical tree crops in Argentina, Brazil, and Paraguay. Its leaves are widely utilized in the preparation of the popular infusion drink “mate.” In Argentina, this tree is cultivated on an area of more than 197,450 ha with a total yield of 718,190 t in 2011/12, representing 85 % of world-wide yerba mate production [1].

In this communication, we describe the first report of a virus identified in yerba mate. The virus family *Endornaviridae* has been recently described encompassing only eight ICTV recognized virus species, and 10 proposed species infecting plants, fungi, and Oomycetes [2, 3]. Endornaviruses are cytoplasmic capsid-less viruses, transmitted vertically to offspring via pollen and ova [4]. Endornaviruses have a double-stranded linear 9.8–17.6 kb RNA genome with a single open reading frame (ORF) encoding a unique ~5,000 aa polyprotein, which is presumed to be processed into functional proteins by virus-encoded proteases [5]. Only a few domains have been described in the ORF, corresponding in several cases to methyl-transferase, helicase, capsular polysaccharide synthesis protein (CPS), UDP-glucose glycosyl-transferase (UGT), and a RNA-dependent RNA polymerase (RdRP), which currently is the only domain found to always be present in an endornavirus [6].

Nucleotide sequence accession number: The genome sequence of the proposed yerba mate endornavirus has been deposited in NCBI GenBank under accession no. KJ634409.

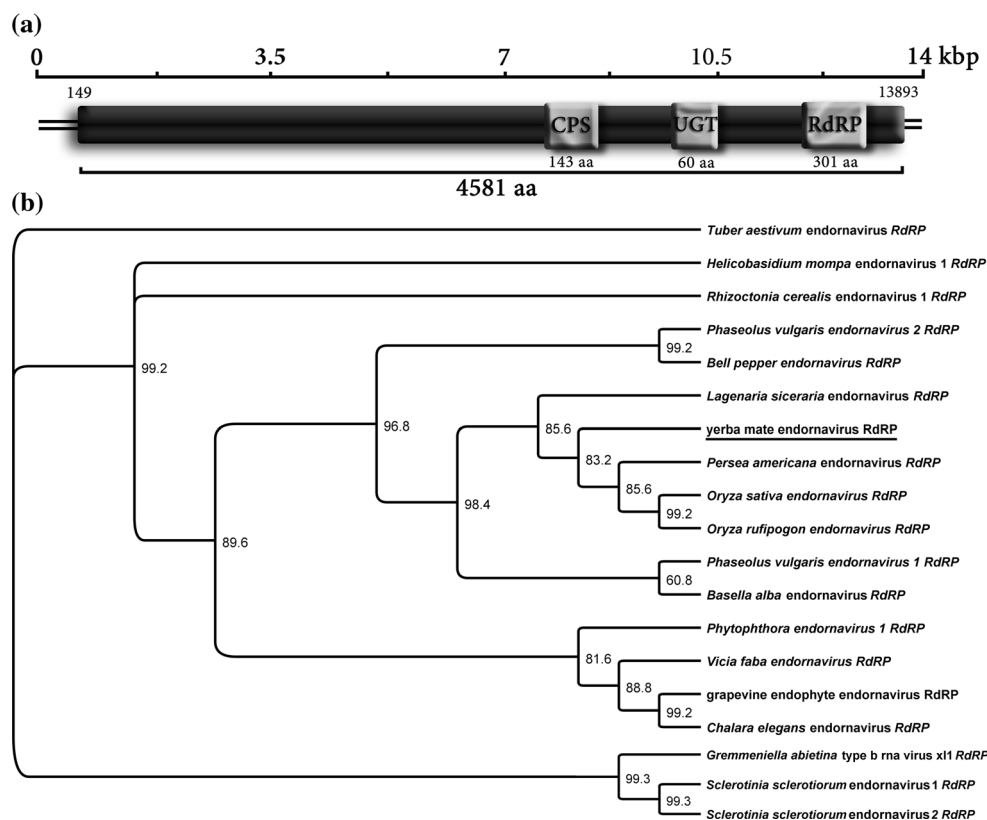
H. J. Debat (✉) · D. A. Ducasse
Instituto de Patología Vegetal, Centro de Investigaciones Agropecuarias, Instituto Nacional de Tecnología Agropecuaria (IPAVE-CIAP-INTA), X5020ICA Córdoba, Argentina
e-mail: hdebat@correo.inta.gov.ar

M. Grabiele · P. M. Aguilera · D. A. Marti
Instituto de Biología Subtropical, Universidad Nacional de Misiones (IBS-UNaM-CONICET), 3300 Posadas, Misiones, Argentina

R. Bubillo
Estación Experimental Agropecuaria Cerro Azul, Instituto Nacional de Tecnología Agropecuaria (EEA Cerro Azul-INTA), 3313 Cerro Azul, Misiones, Argentina

P. D. Zapata
Instituto de Biotecnología Misiones, Facultad de Ciencias Exactas Químicas y Naturales, Universidad Nacional de Misiones (INBIOMIS-FCEQyN-UNaM), 3300 Posadas, Misiones, Argentina

Fig. 1 a The YmEV single linear dsRNA genome is a 13,954-nt long, contains a single 13,743 nt ORF, and presents a 149 nt 5'UTR and a 61 nt 3'UTR. The predicted ORF encodes a 4,581 aa polypeptide. Pfam searches identified a CPS domain between 2,778 and 2,920 aa coordinates, a UGT domain between 3,359 and 3,418 aa, and an RdPR domain between 4,182 and 4,480 aa. **b** Bayesian phylogenetic tree of the RdRP domain of the endornaviruses based on a MUSCLE multiple alignments determined by the Geneious 7.0 platform (Biomatters Ltd.). Numbers at the nodes indicate percentage of bootstrap consensus support values obtained for 1,000 replicates. GenBank accession numbers for the respective viruses are listed in Table 1



Results and discussion

Total leaf RNA from *I. paraguariensis* breeding line Pg538 from INTA EEA-Cerro Azul, Misiones, Argentina was purified and sequenced by the Illumina HiSeqTM-2000 platform. 72,031,388 pair-end 100 nt reads were obtained and assembled by the Trinity method [7] resulting in 44,907 transcripts. The assembled transcriptome was subjected in bulk to homology searches by NCBI-BLASTN. Unequivocally (E value: $6e-160$), an endornavirus species was identified; moreover, low relative sequence similarities suggest it to be a novel species. Iterative mapping of total Illumina reads to the assembled viral genome implies that it accumulates to relatively low levels (0.049 % of total reads). A total of 34,935 mapped 100 nt reads encompass 3,493,500 nt accounting to a $250 \times$ coverage of the virus genome. The presence of the putative endornavirus was confirmed by RT-PCR with specific primers for a predicted RdRP domain encoding region, which resulted in the amplification of an expected ca. 990 nt product when using total RNA of the sequenced yerba mate individual as template. The complete sequence of the proposed yerba mate endornavirus (YmEV) putative genome is estimated to be 13,954-nt long, contains a single 13,743 nt ORF, and presents a 149 nt 5'UTR and a 61 nt 3'UTR (Fig. 1a). The predicted ORF encodes a single 4,581 aa polypeptide.

Searches in *Pfamseq* database version 27.0 with the ORF aa sequence indicate, with low confidence (E value: 0.002), a typical UGT domain between 3,359 and 3,418 aa coordinates and with high confidence a CPS (E value: $2.6e-16$; coordinates 2,778–2,920 aa) and an RdRP domain (E value: $7.8e-25$; coordinates 4,182–4,480 aa). While most endornaviruses have been described to present a helicase domain (HEL), we failed to recognize a putative HEL-like encoding region in YmEV. In agreement with Roossink et al. [2], we suggest that if YmEV harbors a functional HEL domain, it has exceedingly diverged to be recognized by our criteria. As expected from the literature [8–10], the RdRP domain corresponds to the sequence region with the closest homology to the RdRP domain of the proposed *Lagenaria siceraria* endornavirus (LgEV), sharing a 67.1 % similarity at the aa level. The percentage of identity of YmEV domains in relation to every endornavirus described suggests that in general YmEV is more closely related to plant endornavirus that UGT is the second most typical domain of endornaviruses and that while the length of the polypeptide is conserved, global sequence identity is relatively low (Table 1). Multiple MUSCLE amino acid alignments of the RdRP domain, followed by a Bayesian phylogenetic tree generated by the Geneious 7.0 platform (Biomatters Ltd.), indicate that YmEV is more closely and robustly related to the proposed LgEV, *Persea*

Table 1 Amino acid sequence identity percentage of yerba mate endornavirus predicted polypeptide and conserved domains against described endornaviruses

Endornavirus	Host	CPS (%)	UGT (%)	RdRP (%)	Poly (%)	Accession no.
<i>Lagenaria siceraria</i> endornavirus	P	44.4	ND	67.1	36.0	AHK22715.1
<i>Basella alba</i> endornavirus	P	40.6	30.0	66.8	33.4	AB844265.1
<i>Oryza sativa</i> endornavirus	P	32.9	38.3	61.5	34.4	YP_438200.1
<i>Oryza rufipogon</i> endornavirus	P	34.3	40.0	61.8	33.8	YP_438202.1
<i>Persea americana</i> endornavirus	P	44.8	30.0	63.1	38.8	YP_005086952.1
<i>Phaseolus vulgaris</i> endornavirus 1	P	40.6	38.3	62.5	33.9	BAM68539.1
<i>Bell pepper</i> endornavirus	P	ND	14.5	41.9	18.3	BAK52155.1
<i>Phaseolus vulgaris</i> endornavirus 2	P	ND	12.3	42.2	18.0	BAM68540.1
<i>Vicia faba</i> endornavirus	P	ND	ND	39.9	16.6	YP_438201.1
<i>Phytophthora</i> endornavirus 1	O	ND	11.5	42.2	17.9	YP_241110.1
<i>Helicobasidium mompa</i> endornavirus 1	F	ND	11.00	38.0	16.2	YP_003280846.1
<i>Gremmeniella abietina</i> type b ma virus x11	F	ND	ND	25.7	15.2	YP_529670.1
<i>Tuber aestivum</i> endornavirus	F	ND	ND	23.0	13.9	YP_004123950.1
<i>Chalara elegans</i> endornavirus	F	ND	18.5	31.5	17.2	ADN43901.1
grapevine endophyte endornavirus	P	ND	ND	33.2	17.4	YP_007003829.1
<i>Rhizoctonia cerealis</i> endornavirus 1	F	ND	ND	40.9	17.0	YP_008719905.1
<i>Sclerotinia sclerotiorum</i> endornavirus 1	F	ND	ND	23.4	14.4	YP_008169851.1
<i>Sclerotinia sclerotiorum</i> endornavirus 2	F	ND	ND	23.4	14.5	YP_009022070

Poly polypeptide, *CPS* capsular polysaccharide synthase protein, *UGT* UDP-glucose glycosyl-transferase, *RdRP* RNA-dependent RNA polymerase, *ND*. not detected, *P* plant, *F* fungus, *O* Oomycete

americana endornavirus, and *Oryza* endornaviruses (Fig. 1b). The presence of YmEV was not accompanied by any apparent visible viral symptom in the sequenced yerba mate individual. In the future, it would be interesting to explore whether YmEV infection, as most endornaviruses described [3], causes no visible symptoms in *I. paraguariensis*. Due to the intriguing peculiarities of this virus family [2, 3], and the complete lack of the yerba mate virus literature, we consider that the information reported here will be helpful in leading to a new and needed attention to this important topic and crop.

References

1. <http://siia.gob.ar/>
2. M.J. Roossinck, S. Sabanadzovic, R. Okada, R.A. Valverde, J. Gen. Virol. **92**, 2674 (2011)
3. Fukuhara T. and Gibbs M.J. Family Endornaviridae In "Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses". Andrew MQ King, ed. Elsevier, pp519-521 (2012)
4. H. Horiuchi, H. Moriyama, T. Fukuhara, Genes Genet. Syst. **78**, 229 (2003)
5. R. Okada, E. Kiyota, S. Sabanadzovic, H. Moriyama, T. Fukuhara, P. Saha, M.J. Roossinck, A. Severin, R.A. Valverde, J. Gen. Virol. **92**, 2664 (2011)
6. R. Okada, E. Kiyota, H. Moriyama, F. Toshiyuki, R.A. Valverde, Arch. Virol. **159**, 807 (2014)
7. M.G. Grabherr, B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B.W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Nat. Biotechnol. **29**, 644 (2011)
8. F. Villanueva, S. Sabanadzovic, R.A. Valverde, J. Navas-Castillo, J. Virol. **86**, 1282 (2012)
9. D. Song, W.K. Cho, S.-H. Park, Y. Jo, K.-H. Kim, PLoS ONE **8**, e64270 (2013)
10. Y. Espach, H.J. Maree, J.T. Burger, J. Virol. **86**, 13142 (2012)