

Complete genome sequence of a putative new secovirus infecting yam (*Dioscorea*) plants

Evelyn Anly Ishikawa Hayashi¹ · Rosana Blawid² · Fernando Lucas de Melo² · Miguel Souza Andrade² · Gilvan Pio-Ribeiro¹ · Genira Pereira de Andrade¹ · Tatsuya Nagata²

Received: 14 August 2016 / Accepted: 2 October 2016 / Published online: 11 October 2016
© Springer-Verlag Wien 2016

Abstract The complete genome sequence of a new virus infecting yam plants exhibiting mosaic symptom in Brazil was determined. The genome of this virus is composed of two molecules of positive-sense RNAs of 5979 and 3809 nucleotides in length, excluding the poly(A) tails. One large open reading frame (ORF) in each genomic segment (RNA1-ORF1 and RNA2-ORF2) was predicted. The highest amino acid sequence similarity in the Pro-Pol core region of RNA1 and the CP region of RNA2 was observed with chocolate lily virus A (a putative member of the family *Secoviridae*), with 54.6 and 27.7 % identity, respectively. This virus is thus likely to be a new member of the family *Secoviridae*, and we propose the tentative name “dioscorea mosaic-associated virus” (DMaV) for this virus.

Members of the family *Secoviridae* have one or two positive-sense genomic RNA segments. In members with two genome segments, each segment (RNA1 and RNA2) has a genome-linked viral protein (VPg) at its 5' end and a poly(A) tract at a 3' end and encodes a large polyprotein.

Electronic supplementary material The online version of this article (doi:10.1007/s00705-016-3104-4) contains supplementary material, which is available to authorized users.

✉ Tatsuya Nagata
tatsuya@umb.br

¹ Laboratório de Fitovirologia, Departamento de Agronomia, Universidade Federal Rural de Pernambuco, Recife, PE CEP 52171-900, Brazil

² Laboratório de Microscopia Eletrônica, Departamento de Biologia Celular, Universidade de Brasília, Brasília, DF CEP 70910-900, Brazil

The polyprotein of RNA1 contains genes necessary for replication, and the polyprotein of RNA2 contains one or two ORF(s) encoding coat protein(s) (CPs) and a movement protein [1]. The family *Secoviridae* is divided into eight genera (*Comovirus*, *Fabavirus*, *Nepovirus*, *Sequivirus*, *Waikavirus*, *Cheravirus*, *Sadwavirus*, and *Torradovirus*) and includes a few unassigned species [2].

Yam (*Dioscorea* spp.) is an annual herbaceous plant belonging to the Dioscoreaceae family of monocotyledons. Yam plants are commonly propagated vegetatively and are commonly cultivated in Africa, Latin America and Caribbean countries for the consumption of their tubers [3]. Yam production in Brazil is mainly concentrated in the northeast region, where it has significant socio-economic importance [4].

Leaf samples that showed mosaic symptoms were collected from 30 *Dioscorea* plants in the Brazilian states of Pernambuco, Paraíba, and the Federal District (Brasília). To find viral sequences in the samples by high-throughput sequencing, we prepared a virus-rich fraction from the pooled leaf samples from the 30 *Dioscorea* plants and extracted total RNA from this fraction. Briefly, the viruses were partially purified from the pooled samples (2 g of each plant) following the protocol of Cali and Moyer [5] with minor modifications. Total RNA was isolated from the pellet of the semi-purified sample using a ZR Plant RNA MiniPrep kit (Zymo Research, Irvine, USA) following the manufacturer's protocol. The total RNA was sequenced using an Illumina HiSeq 2000 (Illumina Inc. San Diego, USA) with 100-base paired-end reads at Macrogen Inc. (Seoul, South Korea). The reads were organized using CLC Genomics Workbench 6.5 (<http://www.clcbio.com>), and contigs were assembled using Geneious R7.1 (<http://www.geneious.com/>). Contigs were analyzed by BLASTx and protein BLAST searches against the viral reference

genome database (RefSeq) at GenBank. Two of the assembled contigs (5979 bp and 3809 bp in length) were identified as secovirus sequences, and the virus was tentatively named “dioscorea mosaic-associated virus” (DMaV).

cDNA fragments of this isolate of about 500 bp targeting RNA1 were amplified by RT-PCR for the detection of RNA1 in each leaf sample using specific primers (DaMVR1-4719F, 5'-TATCTACAAAATCGTGGAGGAAC-3'; DMaVR1-5020R, 5'-TCAATCTCAGAAAAGGCATGTG-3'). cDNA fragments of expected size were amplified from 20 of the 30 plants used for this study. Nearly full-length genomic cDNA fragments were amplified by reverse transcription polymerase chain reaction (RT-PCR) using a selected isolate from the state of Pernambuco (PE-81) to validate the sequencing results obtained by NGS. cDNA was synthesized again using Superscript III reverse transcriptase (Life Technologies, Grand Island, USA) and oligo-d(T)₅₀M10 (GCAGTGTATCAACGCAGAT₅₀) for both RNA segments. The PCR was performed using the specific primers listed in Supplementary Table 1. These amplified cDNA fragments were sequenced directly at Macrogen Inc. by primer walking. The complete genomic sequences were obtained by amplifying the cDNAs of the 5' terminus of each segment by a modified method of the rapid amplification of cDNA ends (RACE) protocol [6] using the primers in Supplementary Table 1.

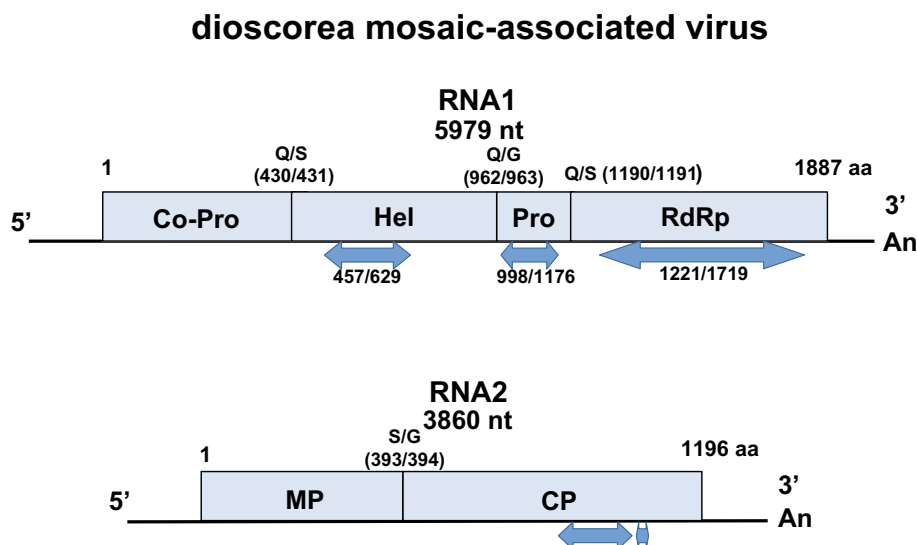
The assembled full-length sequences of RNA1 (KU215538) and RNA2 (KU215539) were 5979 and 3809 nucleotides (nt) in length, respectively, excluding their poly(A) tails. RNA1 contains one large ORF (RNA1-ORF1) of 5663 nt, encoding a predicted polyprotein

(201 kDa) carrying motifs associated with the replication process. Putative cleavage sites in the polyprotein encoded by RNA1 (Fig. 1) were identified at the dipeptides Q/S to cleave a protease co-factor and a helicase, Q/G to cleave the helicase and a protease, and Q/S to cleave the protease and an RNA-dependent RNA polymerase (RdRp). The 5' and 3' untranslated regions (UTRs) of RNA1 were 148 and 151 nt in length, respectively. RNA2 contains one ORF (RNA2-ORF2) of 3590 nt, coding for a predicted polyprotein (134 kDa) with a motif of CP and a cell-to-cell MP separated by a putative S/G cleavage site (Fig. 1). The 5' and 3' UTRs of RNA2 were 72 and 62 nt in length, respectively. The 3' UTRs of RNA1 and RNA2 were 58.4 % identical.

A phylogenetic tree based on the amino acid sequence of the protease-polymerase (Pro-Pol) region between the protease CG motif and the RdRp GDD motif (CG/GDD) of DMaV and other members of the family *Secoviridae* indicated that DMaV is most closely related to chocolate lily virus A (CLVA), a recently reported putative secovirus, and viruses belonging to yet unassigned species in the family *Secoviridae*: black raspberry necrosis virus (BRNV), strawberry mottle virus (SMoV) and strawberry latent ringspot virus (SLRSV) [7–9] (Fig. 2). The genomic organization of DMaV resembled that of secoviruses.

The ICTV species demarcation criteria for the family *Secoviridae* stipulate amino acid sequence identities >80 % in the Pro-Pol region or >75 % in CP [10]. The amino acid sequence identity in the Pro-Pol (CG/GDD) and CP regions of DMaV to other unassigned members of the family *Secoviridae* varied from 54.6 (CLVA) to 42.4 % (SMV) and from 27.7 (CLVA) to 21.9 % (SMoV), respectively, which is below the currently valid species

Fig. 1 Schematic representation of the genomic organization of dioscorea mosaic-associated virus RNA1 and RNA2. Putative cleavage sites are indicated. The motifs are indicated by arrows. Abbreviations: Co-Pro, protease cofactor; Hel, helicase; Pro, protease; RdRp, RNA-dependent RNA polymerase; MP, movement protein; CP, coat protein; Q, glutamine; S, serine; G, glycine. Numbers indicate the amino acid position



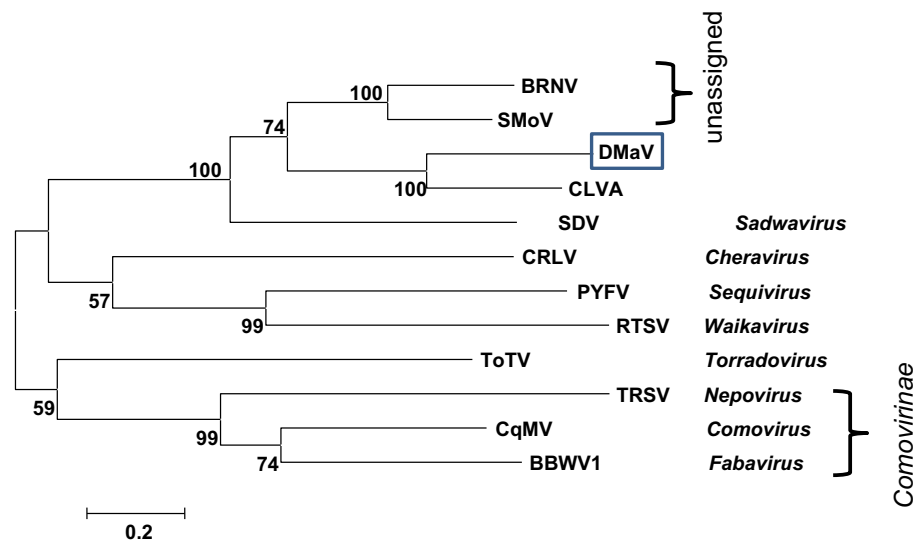


Fig. 2 Phylogenetic relationship of dioscorea mosaic-associated virus with other members of the family *Secoviridae* based on the protease-polymerase core region (CG-GDD motif). A phylogenetic tree was constructed using the maximum-likelihood (ML) method with 500 bootstrap replicates. The following viruses were used in the analysis: BRNV, black raspberry necrosis virus, CCE57809; SMoV,

strawberry mottle virus, NP 599086; CLVA, chocolate lily virus A, YP 004936170; SDV, satsuma dwarf virus, BAA76746; CRLV, cherry rasp leaf virus, CAF21713; PYFV, parsnip yellow fleck virus, BAA03151; RTSV, rice tungro spherical virus, AAA66056; ToTV, tomato torrado virus, ABD38934; CqMV, cowpea mosaic virus, CAA25029; BBWV1, broad bean wilt virus 1, BAD00183

demarcation cutoff. In conclusion, dioscorea mosaic-associated virus is a putative new member of the family *Secoviridae*.

Acknowledgments TN is a research fellow of CNPq (Brazilian Council for Scientific and Technological Development).

Compliance with ethical standards

Conflict of interest The 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

- Sanfaçon H, Wellink J, Le Gall O, Karasev A, van der Vlugt R, Werzel T (2009) Secoviridae: a proposed family of plant viruses within the order Picornavirales that combines the families Sequiviridae and Comoviridae, the unassigned genera Cheravirus and Sadwavirus, and the proposed genus Torradorvirus. *Arch Virol* 154:899–907. doi:10.1007/s00705-009-0367-z
- ICTV Virus Taxonomy 2014. <http://www.ictvonline.org/virusTaxonomy.asp>. Accessed 22 Nov 2015
- Saski CA, Bhattacharjee R, Scheffler BE, Asiedu R (2015) Genomic Resources for Water Yam (*Dioscoreaalata* L.): Analyses of EST-Sequences, De Novo Sequencing and GBS Libraries. *PLoS One*. doi:10.1371/journal.pone.0134031
- Filho F de ACR, Nicolini C, Resende R de O, Andrade GP, Pio-Ribeiro G, Nagata T (2013) The complete genome sequence of a Brazilian isolate of yam mild mosaic virus. *Arch Virol* 158:515–518. doi:10.1007/s00705-012-1509-2
- Cali BB (1981) Purification, serology, and particle morphology of two russet crack strains of sweet potato feathery mottle virus. *Phytopathology* 71:302. doi:10.1094/Phyto-71-302
- Matz M, Shagin D, Bogdanova E, Britanova O, Lukyanov S (1999) Amplification of cDNA ends based on template-switching effect and step-out PCR. *Nucleic Acids Res* 27:1558–1560. doi:10.1093/nar/27.6.1558
- Wylie SJ, Luo H, Li H, Jones MGK (2012) Multiplepolyadenylated RNA viruses detected in pooled cultivated and wild plant samples. *Arch Virol* 157:271–284. doi:10.1007/s00705-011-1166-x
- Lindner JL, Schoen CD, Jelkmann W et al (2002) Characterization and complete nucleotide sequence of Strawberry mottle virus: a tentative member of a new family of bipartite plant picorna-like viruses. *J Gen Virol* 83:229–239. doi:10.1099/0022-1317-83-1-229
- Halgren A, Tzanetakis IE, Martin RR (2007) Identification, characterization, and detection of black raspberry necrosis virus. *Phytopathology* 97:44–50. doi:10.1094/PHYTO-97-0044
- King AMQ, Adams MJ, Lefkowitz EJ (2012) Virus taxonomy: classification and nomenclature of viruses: Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier