

Erratum to: Discovery of two small circular ssDNA viruses associated with the whitefly *Bemisia tabaci*

Erich Y. T. Nakasu¹ · Fernando L. Melo² · Miguel Michereff-Filho¹ ·
Tatsuya Nagata² · Bergmann M. Ribeiro² · Simone G. Ribeiro³ · Cristiano Lacorte³ ·
Alice K. Inoue-Nagata¹

Published online: 22 September 2017
© Springer-Verlag GmbH Austria 2017

Abstract The complete genome sequences of two novel small circular DNA viruses isolated from sweet-potato whiteflies collected in Central-West (AdDF) and Southeast (AdO) regions of Brazil were determined by Next Generation Sequencing (NGS), and confirmed by cloning and Sanger sequencing. The genomes are 2,199 and 2,211 nt-long, respectively, encoding a putative coat protein (CP) and a replication-associated protein (Rep) and showing a genomic organization typical of viruses from the family *Genomoviridae*. Phylogenetic analysis with deduced amino acid sequences of Rep indicates that the virus from AdO is closely related to other members of the genus *Gemycircularvirus*, while the virus from AdDF is related to those of the genus *Gemyduguivirus*. These new genomoviruses are tentatively named bemisia-associated genomovirus AdO and bemisia-associated genomovirus AdDF.

The sweet-potato whitefly (*Bemisia tabaci*) is an insect pest that vectors several plant viruses, particularly those from the genus *Begomovirus*. The worldwide distribution of whiteflies and their virus-associated diseases affect an extensive array of commercial crops, such as soybeans,

cotton, and vegetables [6]. Due to their highly polyphagous behaviour, these insects are consequently exposed to different pathosystems, possibly accumulating a wide range of viruses, including potential novel plant- and insect-infecting species. This premise was confirmed by Rosario et al. [7], who used next generation sequencing (NGS) for assessing viral diversity present in *B. tabaci* whiteflies, an approach commonly termed vector-enabled metagenomics (VEM).

Genomoviruses are single-stranded small circular DNA viruses from the family *Genomoviridae* [3, 9]. Even though the biology of these viruses remains elusive, they were recently described in association with a number of different organisms and environmental samples [3], and classified in nine genera [9]. In the present study, the VEM approach was used to discover novel small circular DNA viruses associated with *B. tabaci* whiteflies.

Two *B. tabaci* sample groups were collected in the Central-West and Southeast regions of Brazil in 2014 and respectively denominated AdDF and AdO. Samples consisted of adult insects feeding on a wide range of plant hosts, such as tomatoes, pumpkins, soybeans, and weeds. Total DNA from a pool of ca. 300 adults from each location was extracted using AllPrep DNA/RNA extraction kit (QIAGEN, Hilden, Germany) and circular DNA was enriched by rolling circle amplification (RCA) [2]. Two different libraries (AdDF and AdO) were prepared and sequenced in an Illumina MiSeq platform at the Catholic University of Brasília (2x250 nt). NGS data were trimmed using Trimmomatic [1] before assembling contigs *de novo* using the Velvet algorithm [10]. Resulting contigs were loaded onto Geneious software (Biomatters, Auckland, New Zealand) and analyzed using BLAST against a RefSeq viral database (downloaded from NCBI on the 5th Oct 2015). Contigs sharing identity with small

The online version of the original article can be found under doi:10.1007/s00705-017-3425-y.

✉ Alice K. Inoue-Nagata
alice.nagata@embrapa.br

- ¹ Embrapa Vegetables, Rodovia BR-060, Km09, Brasília 70275-970, Brazil
- ² Department of Cell Biology, University of Brasília, Brasília 70910-900, Brazil
- ³ Embrapa Genetic Resources and Biotechnology, PqEB, Av. W5 Norte, Brasília 70770-917, Brazil

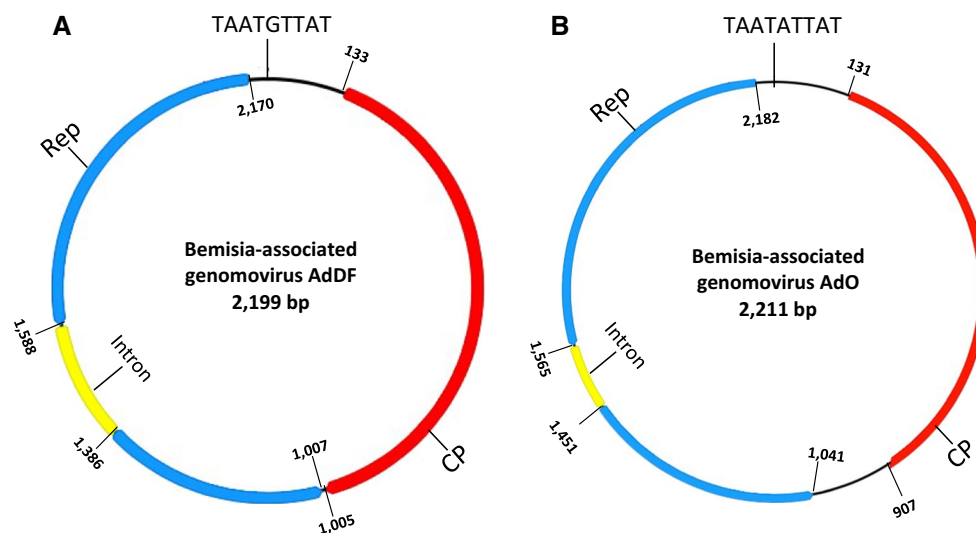
circular DNA viruses were extracted and used as references for extending sequence length using the Geneious mapper. The reads used for mapping were then assembled *de novo* using the Geneious assembler, setting contigs with matching ends to circularize, and producing the complete genome sequence of the putative viruses. To confirm the presence of the viruses in the samples, abutting primers were designed based on the NGS contig sequences. These primers were used to amplify the whole genome from AdDF (F: CTGCTACCGCGGATCTGGACGTTCAAG; R: CTGCTACCGCGGGGGAGTCTTCCAAG) and AdO (F: CTGCTAGAATTCCCGCTCTCAACAACCTTC; R: CTGCTAGAATTCAACATCGTAGTTGCC) using *Taq* Hi-Fi DNA polymerase (Thermo Fisher Scientific, Waltham, USA). The amplicons were cloned into pGEM-T-Easy (Promega, Madison, USA) and sequenced using vector and internal primers (Macrogen Inc., Seoul, South Korea). Putative ORFs were deduced using the ORF finder tool (NCBI) and the putative intron sequence removed from the replication-associated protein (Rep) ORF [8]. Pairwise genetic identity calculations were performed in SDT [5]. Phylogenetic analysis of Rep deduced amino acid sequences was carried out using a MUSCLE alignment of representative genomoviruses, in order to generate a maximum likelihood tree using Mega7 software [4].

Each library yielded a putative complete circular ssDNA virus genome. The first genome, from the AdDF library, was identified from a 1,740 nt contig sharing 74.2% translated amino acid (aa) identity with dragonfly associated gemyduguvirus 1, former dragonfly-associated circular virus 3 (JX185428, tBLASTx, e-value 3.29e-128). This contig was used as a reference for mapping the 2,010,582 reads from the AdDF library. The 47 mapped reads were reassembled, producing a 2,199 nt circular sequence (accession KY230613) with a maximum of 62% genome-wide

nucleotide identity with an isolate of poaceae associated gemycircularvirus 1 (KT253577). This genome contains a slightly modified geminiviral origin of replication TAATGTTAT, and has an intergenic region (IR) comprising 161 nt and two ORFs in opposing directions (Fig. 1). The first ORF (sense) is 873 nt-long and encodes a 290 aa-long putative coat protein (CP) with 85% coverage and 51% aa identity with the CP from dragonfly associated gemyduguvirus 1 (JX185428, 2e-78). The antisense ORF is 1,664 nt-long, with a putative intron of 201 nt, sharing 75% aa identity with the Rep from dragonfly associated gemyduguvirus 1 (YP_009021852, 100% coverage, 4e-180). All typical genomoviridae aa motifs [9] were identified in the predicted aa sequence of this ORF: motif I (LLTYAQ), motif II (THYHA), GRS domain (RVFDIDSYHPNIL-RGI), motif III (YATK), Walker A (GPSRTGKT), Walker B (IFDDM), and motif C (WCNN). Sanger sequencing of three cloned plasmids confirmed the size and sequence of the NGS-derived genome, except for two nucleotide substitutions in one of the sequences. The low genetic identity of the full genome of AdDF to other genomoviruses suggests that it is potentially classified as a new species within *Genomoviridae* [3, 9].

A second viral genome from adult whiteflies (AdO) was assembled from a 680 nt-long contig presenting high aa sequence identity with dragonfly associated gemycircularvirus 1 (JX185429, tBLASTx, 60.4% identity, 4.55e-70). The 2,016 nt circular genomic sequence was assembled from 94 out of 1,749,768 reads from the AdO library, and shares a maximum of 86% nt identity with part of the sequence from pteropus associated gemycircularvirus 3 (KT732797; 40% coverage, e-value 0). This virus genome was then amplified by abutting primers, and a clone, AdO3, was selected and used to compare to the NGS sequence. The sequence of AdO3 was 2,211 nt-long, 195 nt longer

Fig. 1 Schematic view of the genome organization of *Bemisia tabaci* viruses from AdDF (A) and AdO (B) libraries and the putative introns and encoded genes. Figures were generated and modified from Geneious software. CP: coat protein. Rep: replication-associated protein



than the NGS-assembled sequence. This sequence contained an insertion of 27 nt at position 897-923, a second insertion of 167 nt at position 1207-1373, and two 1-nt insertions, besides two nt substitutions. It is speculated that the low coverage of the reads and the paired-end sequence option contributed to the assembly of an incomplete genome. The consensus (KY230614) between the Sanger and NGS sequenced genome was used for further analysis. The full genome sequence shares a maximum of 64% nucleotide identity with bovine associated gemycircularvirus 1, and it should thus be classified as a member of genus *Gemycircularvirus*. This genome encodes two ORFs in opposing directions, has a 158 nt IR, and

the typical genomoviral origin of replication TAATATTAT (Fig. 1). ORF1 is 777 nt-long and encodes a putative CP with 258 aa. This sequence shares 54% aa identity with the CP from pteropus associated gemykolovirus 1 (KT732798, 81% coverage, e-value 1e-67). ORF2 has 1,029 nt with a 113 nt-long intron, and encodes a 342 aa putative Rep protein similar to RepA from pteropus associated gemycircularvirus 3 (KT732797, 99% coverage, 83% identity, e-value 0.0). The predicted aa sequence coded by this ORF also presented all characteristic genomoviridae aa motifs [9]: motif I (LVTYSQ), motif II (LHLHV), GRS domain (DILDVDGRHANVEPSA), motif III (YAIK), Walker A (GGTRTGKT), Walker B (VFDDI), and motif C (WVCN).

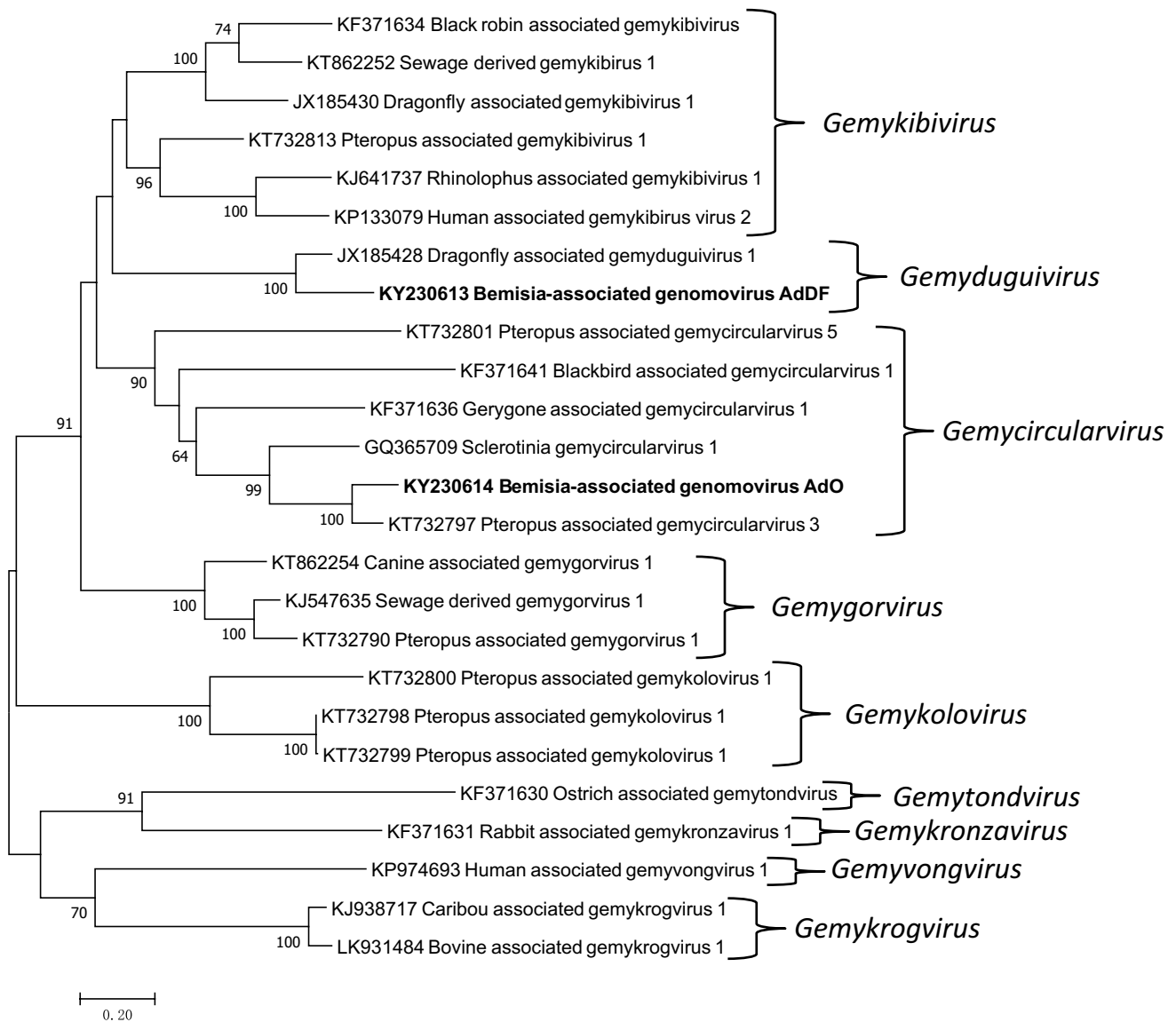


Fig. 2 Maximum likelihood analysis of replication-associated proteins from the new viruses (in bold) and representative sequences of genomoviruses with their new designations, according to the Interna-

tional Committee on Taxonomy of Viruses (ICTV). Genera are indicated according to ICTV. Bootstrap values above 50% are shown at the branch nodes (1,000 replications). Bar: substitutions per site

Phylogenetic analysis of Rep deduced amino acids from the two new viruses was performed after alignment with representative genomovirus sequences (Fig. 2). The AdO sequence is closely related to the majority of gemycircularvirus-like sequences (Fig. 2), including the type-species *Sclerotinia gemycircularvirus 1* (former *Sclerotinia sclerotiorum* hypovirulence associated DNA virus 1). The AdDF sequence clusters with dragonfly associated gemyduguivirus 1 (Fig. 2).

Despite their low identities with other known viruses, the two sequences described here have a typical genome organization of and a genetic relationship with other genomoviruses, indicating they should be considered new members of this family. The virus derived from sample AdO is proposed as bemisia-associated genomovirus AdO, a putative member of the genus *Gemycircularvirus*, whereas the one derived from AdDF is proposed as bemisia-associated genomovirus AdDF, possibly within the genus *Gemyduguivirus*.

Acknowledgements The authors are grateful to Jose Luiz Pereira for assistance in whitefly collection, and Dr. Arvind Varsani for kindly providing invaluable information about genomoviruses and intron splicing procedures, and confirming the Rep sequences. TN, BMR and AKIN are CNPq fellows.

Compliance with ethical standards

This study was funded by CNPq (Grant no. 403414/2013-0) and Embrapa (Grant no. 02.14.00.016.00.00).

Conflict of interest The authors declare that no conflict of interest exists.

References

1. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120
2. Dayaram A, Galatowitsch M, Harding JS, Argüello-Astorga GS, Varsani A (2014) Novel circular DNA viruses identified in *Procordulia grayi* and *Xanthocnemis zealandica* larvae using metagenomic approaches. *Infect Genet Evol* 22:134–141
3. Krupovic M, Ghabrial S, Jiang D, Varsani A (2016) Genomoviridae: a new family of widespread single-stranded DNA viruses. *Arch Virol* 161(9):2633–2643
4. Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33(7):1870–1874
5. Muhire BM, Varsani A, Martin DP (2014) SDT: A virus classification tool based on pairwise sequence alignment and identity calculation. *PLoS One* 9(9):e108277. doi:10.1371/journal.pone.0108277
6. Oliveira MRV, Henneberry TJ, Anderson P (2001) History, current status, and collaborative research projects for *Bemisia tabaci*. *Crop Prot* 20(9):709–723
7. Rosario K, Seah YM, Marr C, Varsani A, Kraberger S, Stainton D, Moriones E, Polston JE, Duffy S, Breitbart M (2015) Vector-enabled metagenomic (VEM) surveys using Whiteflies (Aleyrodidae) Reveal Novel Begomovirus species in the new and old worlds. *Viruses* 7(10):5553–5570
8. Schalk HJ, Matzeit V, Schiller B, Schell J, Gronenborn B (1989) Wheat dwarf virus, a geminivirus of graminaceous plants needs splicing for replication. *EMBO J* 8(2):359–364
9. Varsani A, Krupovic M (2017) Sequence-based taxonomic framework for the classification of uncultured single-stranded DNA viruses of the family *Genomoviridae*. *Virus Evol* 3(1):vew037. doi:10.1093/ve/vew037
10. Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18(5):821–829