JOURNAL OF GENERAL VIROLOGY

RESEARCH ARTICLE

Fontenele *et al.*, *Journal of General Virology* 2021;102:001671

DOI 10.1099/jgv.0.001671

MICROBIOLOGY SOCIETY

Novel circular DNA virus identified in *Opuntia discolor* (*Cactaceae*) that codes for proteins with similarity to those of geminiviruses

Rafaela S. Fontenele[1], Matias Köhler[2], Lucas C. Majure[3], Jesús A. Avalos-Calleros[4], Gerardo R. Argüello-Astorga[4], Fabián Font[5], Andreza H. Vidal[6], Philippe Roumagnac[7,8], Simona Kraberger[1], Darren P. Martin[9], Pierre Lefeuvre[10] and Arvind Varsani[1,11,*]

Abstract

Viral metagenomic studies have enabled the discovery of many unknown viruses and revealed that viral communities are much more diverse and ubiquitous than previously thought. Some viruses have multiple genome components that are encapsidated either in separate virions (multipartite viruses) or in the same virion (segmented viruses). In this study, we identify what is possibly a novel bipartite plant-associated circular single-stranded DNA virus in a wild prickly pear cactus, *Opuntia discolor*, that is endemic to the Chaco ecoregion in South America. Two ~1.8 kb virus-like circular DNA components were recovered, one encoding a replication-associated protein (Rep) and the other a capsid protein (CP). Both of the inferred protein sequences of the Rep and CP are homologous to those encoded by members of the family *Geminiviridae*. These two putatively cognate components each have a nonanucleotide sequence within a likely hairpin structure that is homologous to the origins of rolling-circle replication (RCR), found in diverse circular single-stranded DNA viruses. In addition, the two components share similar putative replication-associated iterative sequences (iterons), which in circular single-stranded DNA viruses are important for Rep binding during the initiation of RCR. Such molecular features provide support for the possible bipartite nature of this virus, which we named utkilio virus (common name of the *Opuntia discolor* in South America) components A and B. In the infectivity assays conducted in *Nicotiana benthamiana* plants, only the A component of utkilio virus, which encodes the Rep protein, was found to move and replicate systemically in *N. benthamiana*. This was not true for component B, for which we did not detect replication, which may have been due to this being a defective molecule or because of the model plants (*N. benthamiana*) used for the infection assays. Future experiments need to be conducted with other plants, including *O. discolor*, to understand more about the biology of these viral components.

INTRODUCTION

The known plant-infecting single-stranded DNA viruses are found in two families – *Geminiviridae* and *Nanoviridae* – both of which have circular genomes. Nanovirus genomes are multipartite, comprising six to eight genome components, while some geminiviruses in the genus *Begomovirus* have bipartite genomes, whereas other geminiviruses have monopartite genomes. In some cases, geminiviruses and nanoviruses can be found in association with sub-viral single-stranded DNA molecules known as satellites. These satellites (alphasatellites, betasatellites and deltasatellites [1–3]) are classified into two families, *Alphasatellitidae* [1]

Received 11 May 2021; Accepted 13 August 2021; Published 02 November 2021

Author affiliations: [1]The Biodesign Center for Fundamental and Applied Microbiomics, Center for Evolution and Medicine, School of Life Sciences, Arizona State University, Tempe, Arizona, 85287, USA; [2]Programa de Pós-Graduação em Botânica, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil; [3]Florida Museum of Natural History, University of Florida, Gainesville, Florida, 32611, USA; [4]División de Biología Molecular, Instituto Potosino de Investigación Científica y Tecnológica, A.C., Camino a la Presa de San José 2055, Lomas 4ta Secc, San Luis Potosi 78216, Mexico; [5]Herbario Museo de Farmacobotánica 'Juan A. Domínguez' (BAF), Universidad de Buenos Aires, Buenos Aires, Argentina; [6]Programa de Pós-Graduação em Biologia Molecular, Universidade de Brasília, Brasília, Brazil; [7]CIRAD, UMR PHIM, 34090 Montpellier, France; [8]PHIM Plant Health Institute, Univ Montpellier, CIRAD, INRAE, Institut Agro, IRD, Montpellier, France; [9]Computational Biology Division, Department of Integrative Biomedical Sciences, Institute of Infectious Diseases and Molecular Medicine, University of Cape Town, Observatory, 7925, Cape Town, South Africa; [10]CIRAD, UMR PVBMT, F-97410 St Pierre, La Réunion, France; [11]Structural Biology Research Unit, Department of Integrative Biomedical Sciences, University of Cape Town, 7925, Cape Town, South Africa.

***Correspondence:** Arvind Varsani, arvind.varsani@asu.edu

Keywords: *Cressdnaviricota*; *Opuntia*; ssDNA virus.

Abbreviations: CFDV, coconut foliar decay virus; CP, capsid protein; CRESS, circular replication-associated protein encoding single stranded; HUH, his-hydrophobe-his; RCR, rolling circle replication; Rep, replication associated protein; SF3, superfamily 3.

One supplementary table and one supplementary data file are available with the online version of this article.

and *Tolecusatellitidae* [4]. While alphasatellites encode their own replication-associated protein and are able to replicate autonomously, betasatellites and deltasatellites do not encode their own replication-associated protein and are therefore dependent on their helper virus for replication. All three groups of satellites are dependent on their helper viruses for movement through the plant, encapsidation and insect transmission. Hence, these satellites have transient interactions with ssDNA viruses, adding to the complexity of multipartite virus infections.

Recently, with the use of viral metagenomic approaches, novel ssDNA viruses have been identified that are not closely related to any of the classified viruses and some of these divergent viruses appear to have putative multipartite genomes. Coconut foliar decay virus (CFDV), for example, is found infecting coconut palms in the Pacific island of Vanuatu and is a ssDNA virus with a putative a tripartite genome [5]. CFDV encodes a capsid protein with highest amino acid similarity to geminiviruses but its virion morphology and genome size are more similar to those of nanoviruses. A tripartite virus with a ssDNA genome was identified infecting a plant pathogenic fungus, *Fusarium graminearum* [6]. This virus, fusarium graminearum gemytripvirus 1 (FgGMTV1), is most closely related to the viruses in the family *Genomoviridae* [7], and is the first known instance of a multipartite ssDNA virus associated with fungi. Other previously identified fungus-infecting viruses have non-segmented or segmented genomes. In addition, circular replication-associated protein encoding single-stranded (CRESS) DNA viruses in the phylum *Cressdnaviricota* [8] with putative multipartite genomes have been identified in honey bees and bat guano samples [9, 10].

Here, we describe a novel circular virus (named utkilio virus) with a likely bipartite genome, i.e. components A and B each ~1.8 kb in size. One component, DNA-A, encodes a replication-associated protein (Rep), and the second, DNA-B, a capsid protein (CP).

METHODS

Sample collection and processing

An *Opuntia discolor* plant was originally collected in Bolivia in 2007 as part of a taxonomic and phylogenetic study of Cactaceae plants in South America. *O. discolor* is a small creeping spreading shrub that is endemic to the ecoregion known as Chaco, which comprises an arid and semi-arid climate with an associated xerophytic biota. Following collection in Bolivia, the plant has been grown at research facilities, initially in Argentina and, since 2017, in Brazil. The *O. discolor* plant was sampled twice in 2019 and total DNA from both collections was extracted using the CTAB method [11].

High-throughput sequencing and genome assembly

The total DNA from the first 2019 samples was used to generate Illumina sequencing libraries and was sequenced on an Illumina HiSeq X platform (paired read 2×150 bp) at RAPiD Genomics (USA). The raw reads were *de novo*-assembled using metaSPAdes v. 3.12.0 [12], and the resulting contigs (>500 nt) were analysed using BLASTX [13] against a GenBank viral RefSeq protein database [14]. Two contigs with similarity to geminiviruses were identified and from these two pairs of abutting primers (FF718_679_F: 5′-TCG TTT TCC ATT TGA AAT GTT TCC TTC ATC-3′, FF718_679_R: 5′-CCA TTT GAA ATA TCC AGT CCA ATT GAA AGC-3′, FF719_669_1_F: 5′-CGT AAT TTC CAT TTC TTA CAA GTG TGA ACG-3′ and FF719_669_1_R: 5′-ATA TTA CCC CTA CTG TAG ACT TTA CTT GGA-3′) were designed to recover the complete sequences of the genome components. Using these primers, PCR was performed with KAPA HiFi HotStart DNA polymerase (KAPA Biosystems, USA) according to the manufacturer's recommended thermal cycling conditions. The amplicons were resolved on a 0.7% agarose gel and the two amplicons with a size of ~2 kb were gel-excised, purified and cloned in the pJET1.2 cloning vector (Thermo Fisher Scientific, USA). The recombinant plasmids were transformed into competent *Escherichia coli* XL1 Blue cells and the recombinant plasmids were Sanger sequenced by primer walking at Macrogen, Inc. (Republic of Korea). The Sanger sequence assembly and annotations were performed using Geneious 11.1.5. (Biomatters Ltd, New Zealand) and SMART (http://smart.embl-heidelberg.de/ [15]). The sequences have been deposited in GenBank under accession numbers MW694325 and MW694325.

Phylogenetic and pairwise identity analyses

Two ~1.8 kb circular sequences were determined, one apparently encoding a replication-associated protein (Rep) and the other a capsid protein (CP) with detectable similarity to homologous proteins encoded by geminiviruses. Two sequence datasets were created with the Rep and CP amino acid sequences encoded by the molecules described in this study along with representative sequences from each geminivirus genus and unclassified geminiviruses (n=54). Each dataset was aligned using MAFFT v.7 [16] and the resulting alignments were used to infer maximum-likelihood (ML) phylogenetic trees with substitution models rtRev+G+I+F for the Rep and rtRev+G+I for the CP, as determined to be best fitting by ProtTest [17]. Approximate likelihood ratio tests (aLRTs) for branch support were performed and branches with <0.8 aLRT support were collapsed using TreeGraph2 [18]. Each ML phylogenetic tree was rooted with Rep and CP amino acid sequences from two members of the family *Genomoviridae* [7].

Infectivity assays

Two copies of each viral component were amplified using specific primers (Table S1, available in the online version of this article), and these were cloned in tandem into the binary vector pJL-89 using the Gibson Assembly method (New England Biolabs, USA), as previously described [19]. Each clone was transformed into competent *E. coli* XL1 Blue cells. To further confirm the correct cloned orientation of the tandem components, clones were analysed by digestion with either HindIII (Utkili virus DNA-A) or ScaI (utkilio virus

DNA-B). Clones containing the two tandemly cloned copies of the utkilio virus DNA-A and DNA-B were used to transform *Rhizobium radiobacter* GV3101.

Infection assays were performed in *Nicotiana benthamiana* plants. In one experiment, 10 *N. benthamiana* plants were inoculated with both utkilio virus DNA-A and DNA-B, and 4 plants were inoculated with either utkilio virus DNA-A or DNA-B on their own. The top leaves of the inoculated plants were sampled at 15 and 45 days post-inoculation (p.i.). Total DNA was extracted from newly emerged leaves at each time point using the GenCatch Plant Genomic DNA Purification kit (Epoch Life Science, USA) according to the manufacturer's instructions. The total DNA was screened by PCR with pairs of abutting primers for each component (FF718_679_F/R and FF719_669_1_F/R) using KAPA HiFi HotStart DNA polymerase (KAPA Biosystems, USA). The amplified fragments were cloned and Sanger sequenced.

RESULTS AND DISCUSSION

Identification and recovery of geminivirus-like sequences

An *O. discolor* plant was initially collected in Bolivia in 2007 as part of a phylogenetic study of the family Cactaceae. Prior to 2019 when DNA was extracted from the plant for high-throughput sequencing, it was maintained at a research facility in Argentina until 2017 and thereafter at a facility in Brazil. After *de novo* assembly of the Illumina sequencing reads and BLASTX searches, two contigs of 1806 and 1816 nt that potentially encode proteins with similarity to geminivirus proteins were identified. As the total DNA extracted from the *O. discolor* sample used for Illumina sequencing was no longer available, a second plant sample was obtained in 2019 and used for further analyses. Two pairs of abutting primers designed from the *de novo*-assembled HTS contigs were used to recover the circular geminivirus-like sequences.

Sequence analysis of geminivirus-like sequences

The amplicons obtained corresponded to two sequences of ~1.8 kb. One sequence was 1852 nt in size, having only one large open reading frame (ORF) encoding a putative protein of 350 amino acids with highest amino acid sequence identity of 51.5% to the replication-associated protein (Rep) encoded by the geminivirus, Opuntia virus 2 [20] (Fig. 1, Data S1). The second sequence had 1862 nt and one large ORF encoding a 323 amino acid protein with the highest amino acid sequence identity of 30.5% to the capsid protein (CP) of the geminivirus Exomis microphylla-associated virus [21] (Fig. 1, Data S1). Two other ORFs, named V2 and V3, that could potentially express proteins of 93 and 133 amino acids, were identified upstream of the CP coding ORF, but have no detectable homology with other described proteins (Fig. 1). However, both of the proteins putatively encoded by these two ORFs contained some conserved domains identified using SMART [15] that have been found within other geminivirus proteins. The V3 protein has a 22 amino acid transmembrane domain, which has been associated with putative movement proteins encoded by geminiviruses [22–24]. The V2 protein contains a 54 amino acid zinc-finger3 CxxC domain, other zinc-finger domains have been associated with the geminivirus transcriptional-activator protein (TrAP/AC2) post-transcriptional gene silencing functions [25, 26], although the V2 protein described here has no detectable homology to TrAP.

Geminiviruses replicate via a rolling-circle mechanism [27, 28], with the virus-encoded Rep recognizing specific replication-associated iterative sequences ('iterons') [29] and a nonanucleotide sequence present in a hairpin structure at the virion strand origin of replication (*v-ori*). In addition, in the case of bipartite begomoviruses, the two genomic components share a common region of ~200 nt containing iteron sequences where the *v-ori* hairpin structure is located. These 'common regions' (CRs) enable the components to be replicated by the same Rep, which, in the case of all known multipartite ssDNA viruses, is encoded on one of the components. Therefore, we sought to identify whether any CR-like features were present in the two identified sequences that would further suggest that they are two components of the same virus.

Comparing the intergenic region of both virus components, we identified two possible CRs, hereafter referred to as CR1 and CR2 (Fig. 1). CR2 is 223 nt long and shares 99% nucleotide sequence identity between both molecules, and there is also a putative hairpin structure containing a possible *v-ori* nonanucleotide motif 'TACACCTTA'. However, this hairpin structure is smaller than what is commonly described for geminiviruses, and CR2 does not seem to contain any putative iteron sequences. On the other hand, CR1 (335 nt long) contains a putative hairpin structure that is more similar to those of known geminiviruses with a nonanucleotide sequence 'TAACATTGC', TATA boxes and apparent iteron sequences shared by both components. The iteron sequence, 'GGTGAC', occurs as three repeats in both molecules, with one of the molecules having an additional inverted repeat near the nonanucleotide sequence (Fig. 1). Despite the CR1 sequences sharing only ~60% identity between the two components, based on the important features required for initiation of replication, we expect that CR1 is where the origin of replication is located.

Due to the shared molecular signals likely associated with the origin of replication, we believe that these two molecules (DNA-A and DNA-B) are putative components of a multipartite virus, which we have named utkilio virus. Utkilio is a vernacular Bolivian name of *O. discolor*. Although no other viral-like sequences were identified in the *de novo*-assembled contigs, it is plausible that other molecules might be additional components of, or be associated with, utkilio virus.

The Rep protein encoded by utkilio virus DNA-A shares 31.4–51.5% amino acid sequence identity to the Reps encoded by geminiviruses (Data S1). Another important feature that provides some evidence for the Rep function of the utkilio virus is the presence of conserved motifs shared not only by the Reps of geminiviruses but between all rep-encoding

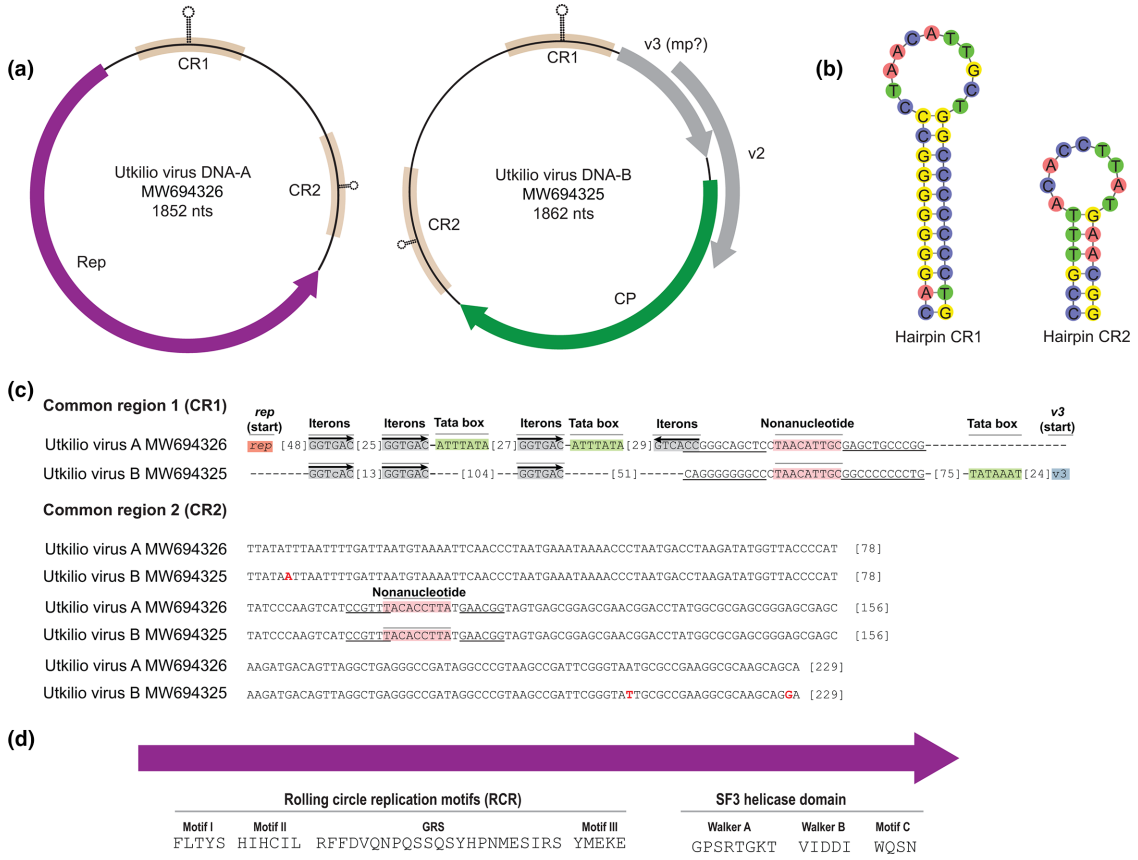

Fig. 1. (a) Genomic organization of the utkilio virus components A and B showing the encoded proteins replication-associated protein (Rep), capsid protein (CP), V2 and V3, as well as the common region 1 and 2 (CR1 and CR2) with hairpin structure. (b) Hairpin structures contained within the CR1 and CR2. (c) Comparison of the CR1 and CR2 between the utkilio virus components A and B. The relative positions of the 'iteron' sequences, the TATA box and the nonanucleotide sequence within the hairpin are depicted over the CR1 sequences. In the CR2 alignment, differences in the nucleotides sequences are marked in red and the nonanucleotide within the hairpin is highlighted. (d) Rolling-circle replication motifs and helicase domain motifs contained within the Rep encoded by utkilio virus component A.

ssDNA viruses. The utkilio virus Rep contains the HUH (his-hydrophobe-his) endonuclease domain [30, 31] and the superfamily 3 (SF3) helicase domain [32, 33]. The HUH domain contains conserved motifs I, II and III as well as a geminivirus-specific GRS motif [34], all of which are important for rolling-circle replication (RCR), while the SF3 domain contains the motifs Walker A, Walker B and motif C (Fig. 1).

ML phylogenetic analysis of the utkilio virus Rep protein together with those of other geminiviruses shows that the sequence is basal to a cluster of sequences of the genera *Begomovirus*, *Curtovirus*, *Eragrovirus*, *Grablovirus*, *Opunvirus*, *Topocuvirus* and *Turncurtovirus*, and the currently unclassified viruses, Opuntia virus 2 [20], Polygala garcinii-associated virus [21] and parsley yellow leaf curl virus [35] (Fig. 2).

The CP encoded by the utkilio virus DNA-B shares 19.9–30.5% amino acid identity with those encoded by geminiviruses (Data S1). ML phylogenetic analysis of the encoded CP together with those of other geminiviruses indicated that utkilio virus CP sits basal to a cluster containing CP sequences of begomoviruses, and the recently established new genera *Citlodavirus* and *Mulcrilevirus* [36, 37], as well as the Polygala garcinii-associated virus [21].

Infectivity experiments

The shared molecular features in utkilio virus DNA-A and DNA-B suggest that they are putative components of a bipartite virus. To test the infectivity of DNA-A and DNA-B, we experimentally inoculated *N. benthamiana* plants either individually or in combination. As expected, the *N. benthamiana* plants inoculated with only utkilio virus DNA-B displayed no evidence of infection in newly emerged leaves. This was expected because replication and systemic spread of DNA-B would require the Rep encoded by DNA-A to initiate RCR. On the other hand, three out of four *N. benthamiana* plants inoculated with only the Rep encoding utkilio virus DNA-A displayed evidence of the DNA-A molecule spreading systemically, in that DNA-A was detectable by PCR in newly emerged leaves at 45 days p.i. There were no perceptible visual symptoms in the plants inoculated with utkilio virus DNA-A.

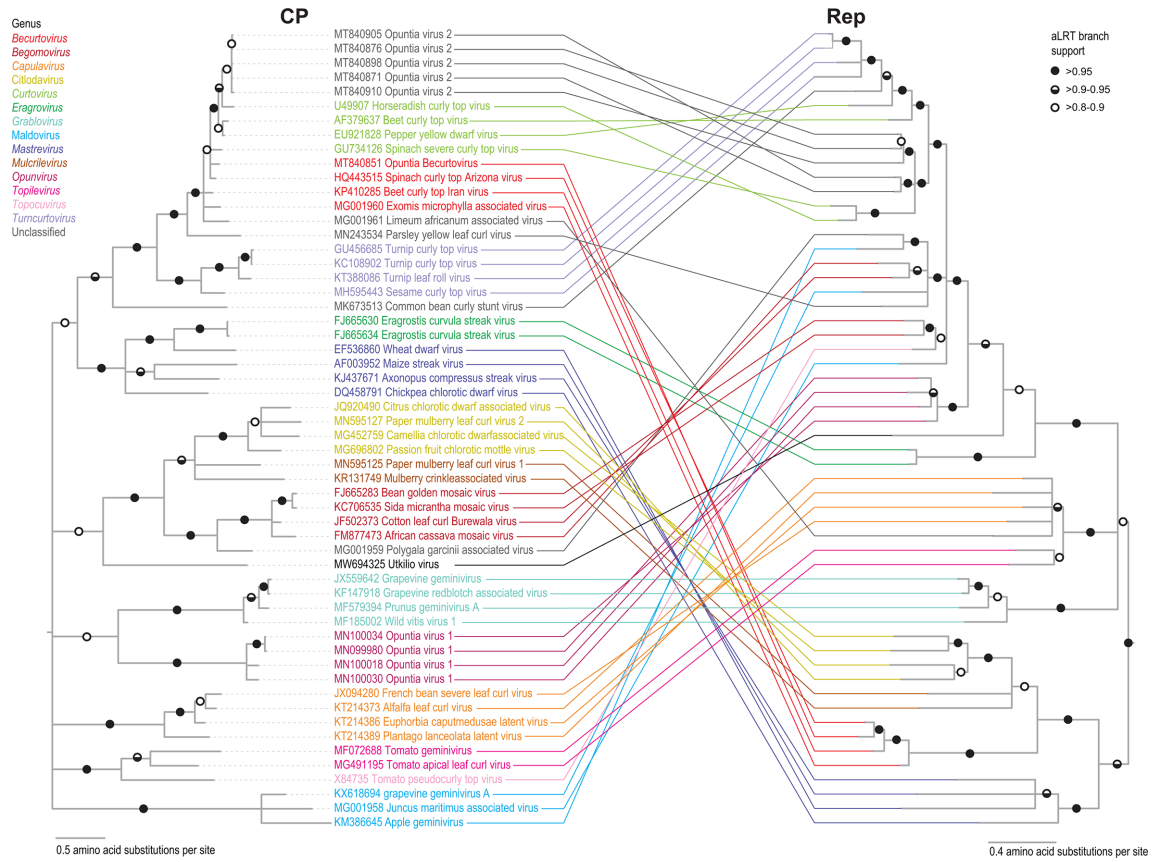

Fig. 2. Maximum-likelihood phylogenetic tree of the Rep and CP amino acid sequences of utkilio virus components and representative sequences from various genera in the family *Geminiviridae*. Branches with <0.8 aLRT support have been collapsed and the trees are rooted with sequences of genomoviruses [7].

When *N. benthamiana* plants were co-inoculated with both utkilio virus DNA-A and DNA-B components, only 1 out of 12 plants showed evidence of systemic spread of the DNA-A at 15 days p.i. However, at 45 days p.i., 9 out of the 12 plants displayed PCR evidence of systemic DNA-A spread. The DNA-B was not detected in any of the newly emerged leaves of co-inoculated plants at either 15 or 45 days p.i. Again, no perceptible visual symptoms were identified in the co-inoculated plants.

Using 5 µg of total DNA from *N. benthamiana* plants infected with DNA-A, we were unable to visualize the virus replicative forms via Southern blot analysis, which indicates that the DNA-A accumulated at very low titres. This was the case for both infectivity assays where DNA-A was detected (with or without co-infection with DNA-B).

It is important to emphasize that *N. benthamiana* is not the original host of utkilio virus: a fact that may explain our inability to detect DNA-B in newly emerged leaves. Moreover, we recovered two DNA-B clones from the original *O. discolor* plant, but both clones only shared 99% sequence identity with four single-nucleotide polymorphisms that did not affect the size of the ORFs, but there are some variations in the amino acid sequences. We cannot rule out that the DNA-B used in the infectivity experiments could be defective. It is also possible that there might be additional utkilio virus components that remain unidentified and that these might impact on infection dynamics. Despite screening the *de novo*-assembled contigs using the intergenic region of both components, we were unable to identify any other putative components. Future experiments should be conducted to identify the host range of this virus and, most importantly, test the infectivity of the DNA-A and DNA-B pair in cactus plants.

CONCLUDING REMARKS

Here we report the identification and characterization of a novel putatively bipartite circular single-stranded DNA virus identified in *O. discolor* that encodes Rep and CP that are homologous to those encoded by members of the family *Geminiviridae*. As the analysed *O. discolor* plant was originally collected in Bolivia but grown in Argentina for 10 years and later in Brazil for a further 3 years, it is not possible to determine which country the infection occurred in and it is therefore also not possible to associate the virus with a precise geographical origin. Future collection of *O. discolor* and neighbouring plants in those areas will be essential to determine the origin of the virus and possible host range.

The degree of similarity of utkilio virus-encoded protein sequences to those of geminiviruses and the molecular features its apparent genome components share with multipartite begomoviruses and nanoviruses indicate the putative multipartite nature of this new virus. Nonetheless, this virus is evolutionarily distinct from geminiviruses and future research to determine possible insect vectors and virion particles morphology is essential.

We sought to further investigate this with infectivity assays. The Rep-encoding utkilio virus DNA-A systemically infected *N. benthamiana* plants in the presence of utkilio virus DNA-B but also when inoculated individually. The utkilio virus DNA-B, however, was not found to systemically infect inoculated plants in any of the experiments conducted. It is unlikely that utkilio virus is adapted to *N. benthamiana*, and future experiments need to be conducted in its original host species. It is interesting, nonetheless, that utkilio virus DNA-A was able to spread systemically and replicate within *N. benthamiana*. It has been documented for some bipartite begomoviruses that DNA-A components are able to replicate and infect *N. benthamiana* systematically without the presence of movement protein-encoding DNA-B components [38–45]. In fact, an experimental study showed that the DNA-A of the begomovirus abutilon mosaic virus can spread through the plant and enter the cell nucleus for replication in an independent way, even in the presence of DNA-B [43]. Although the infectivity experiments with utkilio virus are consistent with those for begomoviruses, they should not be considered to be definitive, and more experimental work is necessary to determine the role of the DNA-B in infections. It is possible that the utkilio virus DNA-B may be replicating, but at lower levels, and is not able to move systematically and/or is not well adapted to do so in *N. benthamiana*, or that the cloned DNA-B was defective In addition, it cannot be ruled out that there could be other genome components associated with utkilio virus in nature that we were not able to detect.

Another important point to keep in mind is the possibility that these segments may have arisen in a long-term *O. discolor* infection due to a genome reduction process. This process can occur when a virus infects a 'dead-end host' and is not able to be transmitted. The virus is then trapped in the host and, if selection favours faster genome replication over all else, genome reduction might occur as a consequence of this. This is another hypothesis that needs to be further investigated.

Funding information
The molecular work was supported by a start-up grant awarded to A.V. from Arizona State University, USA. Financial support for this project was supported in part through start-up funds to L.C.M from the University of Florida, USA.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. Briddon RW, Martin DP, Roumagnac P, Navas-Castillo J, Fiallo-Olivé E, *et al*. Alphasatellitidae: A new family with two subfamilies for the classification of geminivirus- and nanovirus-associated alphasatellites. *Arch Virol* 2018;163:2587–2600.
2. Fiallo-Olive E, Tovar R, Navas-Castillo J. Deciphering the biology of deltasatellites from the New World: maintenance by New World begomoviruses and whitefly transmission. *New Phytol* 2016;212:680–692.
3. Gnanasekaran P, Kishore Kumar R, Bhattacharyya D, Vinoth Kumar R, Chakraborty S. Multifaceted role of geminivirus associated betasatellite in pathogenesis. *Mol Plant Pathol* 2019;20:1019–1033.
4. Adams MJ, Lefkowitz EJ, King AMQ, Harrach B, Harrison RL, *et al*. Changes to taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2017). *Arch Virol* 2017;162:2505–2538.
5. Gronenborn B, Randles JW, Knierim D, Barrière Q, Vetten HJ, *et al*. Analysis of DNAs associated with coconut foliar decay disease implicates a unique single-stranded dna virus representing a new taxon. *Sci Rep* 2018;8:5698.
6. Li P, Wang S, Zhang L, Qiu D, Zhou X, *et al*. A tripartite ssDNA mycovirus from a plant pathogenic fungus is infectious as cloned DNA and purified virions. *Sci Adv* 2020;6:eaay9634.
7. Varsani A, Krupovic M. Sequence-based taxonomic framework for the classification of uncultured single-stranded DNA viruses of the family Genomoviridae. *Virus Evol* 2017;3:vew037.
8. Krupovic M, Varsani A, Kazlauskas D, Breitbart M, Delwart E, *et al*. Cressdnaviricota: a virus phylum unifying seven families of rep-encoding viruses with single-stranded, circular DNA genomes. *J Virol* 2020;94:e00582-20.
9. Kraberger S, Cook CN, Schmidlin K, Fontenele RS, Bautista J, *et al*. Diverse single-stranded DNA viruses associated with honey bees (Apis mellifera). *Infect Genet Evol* 2019;71:179–188.
10. Male MF, Kraberger S, Stainton D, Kami V, Varsani A. Cycloviruses, gemycircularviruses and other novel replication-associated protein encoding circular viruses in Pacific flying fox (Pteropus tonganus) faeces. *Infect Genet Evol* 2016;39:279–292.
11. Doyle JJ, Doyle JL. Isolation ofplant DNA from fresh tissue. *Focus* 1990;12:39–40.
12. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, *et al*. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19:455–477.
13. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–410.
14. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, *et al*. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* 2016;44:D733-45.
15. Letunic I, Doerks T, Bork P. SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res* 2012;40:D302-305.
16. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;30:772–780.
17. Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 2011;27:1164–1165.
18. Stover BC, Muller KF. TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinformatics* 2010;11:7.
19. Ferro MMM, Ramos-Sobrinho R, Xavier CAD, Zerbini FM, Lima GSA, *et al*. New approach for the construction of infectious clones of a circular DNA plant virus using Gibson Assembly. *J Virol Methods* 2019;263:20–23.
20. Fontenele RS, Salywon AM, Majure LC, Cobb IN, Bhaskara A, *et al*. New world cactaceae plants harbor diverse geminiviruses. *Viruses* 2021;13.
21. Claverie S, Bernardo P, Kraberger S, Hartnady P, Lefeuvre P, *et al*. From spatial metagenomics to molecular characterization of plant viruses: a geminivirus case study. *Adv Virus Res* 2018;101:55–83.
22. Hehnle S, Wege C, Jeske H. Interaction of DNA with the movement proteins of geminiviruses revisited. *J Virol* 2004;78:7698–7706.

23. Boulton MI. Functions and interactions of mastrevirus gene products. *Physiol Mol Plant Pathol* 2002;60:243–255.
24. Bernardo P, Golden M, Akram M, Naimuddin N, Nadarajan N, *et al*. Identification and characterisation of a highly divergent geminivirus: evolutionary and taxonomic implications. *Virus Res* 2013;177:35–45.
25. Van Wezel R, Liu H, Wu Z, Stanley J, Hong Y. Contribution of the zinc finger to zinc and DNA binding by a suppressor of posttranscriptional gene silencing. *J Virol* 2003;77:696–700.
26. van WR, Dong X, Liu H, Tien P, Stanley J, *et al*. Mutation of three cysteine residues in Tomato yellow leaf curl virus-China C2 protein causes dysfunction in pathogenesis and posttranscriptional gene-silencing suppression. *Mol Plant Microbe Interact* 2002;15:203–208.
27. Jeske H, Lutgemeier M, Preiss W. DNA forms indicate rolling circle and recombination-dependent replication of Abutilon mosaic virus. *EMBO J* 2001;20:6158–6167.
28. Hanley-Bowdoin L, Bejarano ER, Robertson D, Mansoor S. Geminiviruses: masters at redirecting and reprogramming plant processes. *Nat Rev Microbiol* 2013;11:777–788.
29. Arguello-Astorga GR, Ruiz-Medrano R. An iteron-related domain is associated to Motif 1 in the replication proteins of geminiviruses: identification of potential interacting amino acid-base pairs by a comparative approach. *Arch Virol* 2001;146:1465–1485.
30. Koonin EV. A common set of conserved motifs in a vast variety of putative nucleic acid-dependent ATPases including MCM proteins involved in the initiation of eukaryotic DNA replication. *Nucleic Acids Res* 1993;21:2541–2547.
31. Koonin EV, Ilyina TV. Computer-assisted dissection of rolling circle DNA replication. *Biosystems* 1993;30:241–268.
32. Gorbalenya AE, Koonin EV. Helicases: amino acid sequence comparisons and structure-function relationships. *Curr Opin Struct Bio* 1993;3:419–429.
33. Gorbalenya AE, Koonin EV, Wolf YI. A new superfamily of putative NTP-binding domains encoded by genomes of small DNA and RNA viruses. *FEBS Lett* 1990;262:145–148.
34. Nash TE, Dallas MB, Reyes MI, Buhrman GK, Ascencio-Ibañez JT, *et al*. Functional analysis of a novel motif conserved across Geminivirus rep proteins. *J Virol* 2011;85:1182–1192.
35. Hasanvand V, Heydanejad J, Massumi H, Kleinow T, Jeske H, *et al*. Isolation and characterization of a novel geminivirus from parsley. *Virus Res* 2020;286:198056.
36. Roumagnac P, Varsani A, Martin DP, Lett J-. M. Approved Proposals - Plant Viruses; 2020. https://talk.ictvonline.org/files/ictv_official_taxonomy_updates_since_the_8th_report/m/plant-official/11081 [accessed 1 March 2021].
37. Walker PJ, Siddell SG, Lefkowitz EJ, Mushegian AR, Adriaenssens EM, *et al*. Changes to virus taxonomy and to the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2021). *Arch Virol* 2021;166:2633–2648.
38. Chakraborty S, Pandey PK, Banerjee MK, Kalloo G, Fauquet CM. Tomato leaf curl Gujarat virus, a New Begomovirus Species Causing a Severe Leaf Curl Disease of Tomato in Varanasi, India. *Phytopathology* 2003;93:1485–1495.
39. Galvao RM, Mariano AC, Luz DF, Alfenas PF, Andrade EC, *et al*. A naturally occurring recombinant DNA-A of a typical bipartite begomovirus does not require the cognate DNA-B to infect Nicotiana benthamiana systemically. *J Gen Virol* 2003;84:715–726.
40. Evans D, Jeske H. DNA B facilitates, but is not essential for, the spread of abutilon mosaic virus in agroinoculated Nicotiana benthamiana. *Virology* 1993;194:752–757.
41. Hou Y-M, Paplomatas EJ, Gilbertson RL. Host adaptation and replication properties of two bipartite geminiviruses and their pseudorecombinants. *MPMI* 1998;11:208–217.
42. Klinkenberg FA, Stanley J. Encapsidation and spread of African cassava mosaic virus DNA A in the absence of DNA B when agroinoculated to Nicotiana benthamiana. *J Gen Virol* 1990;71:1409–1412.
43. Weigel K, Pohl JO, Wege C, Jeske H. A population genetics perspective on geminivirus infection. *J Virol* 2015;89:11926–11934.
44. Saunders K, Salim N, Mali VR, Malathi VG, Briddon R, *et al*. Characterisation of Sri Lankan cassava mosaic virus and Indian cassava mosaic virus: evidence for acquisition of a DNA B component by a monopartite begomovirus. *Virology* 2002;293:63–74.
45. Rochester DE, Kositratana W, Beachy RN. Systemic movement and symptom production following agroinoculation with a single DNA of tomato yellow leaf curl geminivirus (Thailand). *Virology* 1990;178:520–526.