1 **A glimpse into the DNA virome of the unique “living fossil” *Welwitschia mirabilis***

2

3 Humberto Debat^{1,2}, Nicolás Bejerman^{1,2}

4

5 ¹ Instituto de Patología Vegetal – Centro de Investigaciones Agropecuarias – Instituto Nacional de
6 Tecnología Agropecuaria (IPAVE-CIAP-INTA), Camino 60 Cuadras Km 5,5 (X5020ICA), Córdoba,
7 Argentina

8 ² Consejo Nacional de Investigaciones Científicas y Técnicas. Unidad de Fitopatología y Modelización
9 Agrícola

10

11

12 Corresponding authors: Humberto Debat, debat.humberto@inta.gob.ar;

13 Nicolás Bejerman, bejerman.nicolas@inta.gob.ar

14

15

16 Abstract

17 Here, we report the identification and characterization of four novel DNA viruses associated with
18 *Welwitschia mirabilis*. Complete circular virus-like sequences with affinity to *Caulimoviridae* members
19 and geminiviruses were detected and characterized from *Welwitschia mirabilis* genomic data. The two
20 newly *Caulimoviridae*-like viruses have been tentatively named as *Welwitschia mirabilis* virus 1 and 2
21 (WMV1-WMV2); whereas the two identified geminiviruses were named as *Welwitschia mirabilis*
22 associated geminivirus A and B (WMA-GVA-WMA-GVB). Phylogenetic analysis suggests that WMV1-2
23 belong to a proposed genus of *Caulimoviridae*-infecting gymnosperms. WMA-GVA-B are
24 phylogenetically related with both mastreviruses and capulaviruses and likely represent a distinct
25 evolutionary lineage within geminiviruses.

26
27 Key words: welwitschia; caulimovirid, geminivirus; data mining; DNA virome.

28
29 *Welwitschia mirabilis* is a distinctive long-living dwarf tree with only two leaves growing throughout the
30 entire plant life [1]. This plant is a monotypic gymnosperm species that is the lone member of the family
31 *Welwitschiaceae*, which is endemic to the Namib Desert in both northern Namibia and southern Angola.
32 *Welwitschia* is usually referred as a “living fossil” because is one of the longest-lived plants on Earth,
33 living up to 3,000 years old [2] and there is a fossil record of this species of 112 My, phylogenetically
34 isolated within the order Gnetales [3]. This exotic and enigmatic plant has been thoroughly studied with a
35 focus in its biological, physiological and ecological peculiarities. Nevertheless, regarding plant health, as
36 of today there has been only one record of a virus linked to this ancient plant: the potyvirus catharanthus
37 mosaic virus (CatMV) which was detected in a *Welwitschia* plant grown ex situ in a domestic garden in
38 western Australia [4].

39 A vast number of novel viruses, many of them not inducing any apparent symptoms, have been identified
40 using metagenomic approaches, which revealed our constrained knowledge about the richness and
41 diversity of the plant virosphere [5-7]. In the context of an advancement in the implementation of high-
42 throughput sequencing technologies as a steady tool in virus discovery, it has been stated that those newly
43 identified viruses that are known only from metagenomic data can, should, and have been incorporated
44 into the official classification scheme of the International Committee on Taxonomy of Viruses (ICTV) [8,
45 9]. In the recent years, the analysis of public data constitutes an emerging source of novel *bona fide* plant
46 viruses, which allows the reliable identification of new viruses in hosts with no previous record of virus
47 infections [10, 11]. In general, plant viruses reported to date based in metagenomic data have been
48 detected analyzing publically available plant transcriptome datasets, resulting mostly in the detection of
49 RNA viruses which could have been inadvertently co-isolated with host RNAs during library preparation
50 [10, 12-15]. WGS DNA datasets have received less attention in virus discovery studies, perhaps given its
51 higher complexity in terms of database size and technical difficulties during processing. However, given
52 their widespread transcription or RNA intermediates it is also possible to detect signatures of DNA
53 viruses from RNAseq datasets, in some cases allowing its characterization and the assembly of complete
54 virus genomes [6, 16-18].

55 The members belonging to the family *Caulimoviridae* have a double stranded DNA (dsDNA) circular
56 genome with a size ranging between 7.1 kb to 9.8 kb including one to eight open reading frames (ORFs)
57 [19] and are the only plant-infecting viruses known to replicate by reverse transcription [20]. The
58 *Geminiviridae* family is the most numerous family of plant-infecting viruses and its members have
59 genomes comprising one or two single-stranded, circular DNAs (ssDNA) [21].

60 Here, we analyzed transcriptomic and WGS DNA datasets corresponding to *Welwitschia mirabilis*,
61 publicly available at the NCBI SRA database, which resulted in the identification and assembly of the
62 complete genomic sequences of four novel DNA viruses, the first associated with this “fossil” plant. Two
63 of them are related to *Caulimoviridae* members and were tentatively named as *Welwitschia mirabilis*
64 virus 1 (WMV1), *Welwitschia mirabilis* virus 2 (WMV2); whereas the other two novel viruses are related
65 with the geminiviruses and were named as *Welwitschia mirabilis* associated geminivirus A (WMaGVA)
66 and *Welwitschia mirabilis* associated geminivirus B (WMaGVB).

67 The raw data analyzed in this study corresponds to 30 publicly available *Welwitschia mirabilis* RNAseq
68 and WGS DNA NGS libraries (Supplementary Table 1) corresponding to a variety of samples from wild
69 (Namib desert) and ex situ greenhouse plants (China and USA), of different developmental stages (young
70 and old leaves), tissues (meristematic and peripheral meristematic), organs (cones, roots, leaves), from
71 male and female individuals. The RNAseq raw reads from each library were pre-processed by trimming
72 and filtering with the Trimmomatic tool as implemented in
73 <http://www.usadellab.org/cms/?page=trimmomatic>, the resulting reads were assembled *de novo* with
74 Trinity v2.6.6 release with standard parameters. The WGS DNA dataset was assembled using Unicycler
75 (Version 0.4.8) with standard parameters as implemented in the Galaxy platform <https://usegalaxy.org/>.
76 *De novo* assembled sequences were subjected to bulk local BLASTX searches (E-value < 1e-5) against a
77 refseq virus protein database available at
78 <ftp://ftp.ncbi.nlm.nih.gov/refseq/release/viral/viral.1.protein.faa.gz>. The resulting hits were clustered
79 using the CD-HIT tool available at <http://weizhong-lab.ucsd.edu/cdhit-web-server/cgi-bin/index.cgi> with
80 a sequence identity cut-off of 0.95. The resulting contigs were explored by hand. Our attention was
81 caught on that our best (but substantially low identity) hits corresponded to DNA viruses, more
82 specifically two ca. 2.9 kb sequences from the SRA SRX9189731 corresponding to DNA-Seq of
83 *Welwitschia* young leaves from a green house in China [22] obtained best significant hits (E-value = 1e-
84 45 and 5e-48, identity = 53.54% and 40.22%) with the replication associated protein of the geminiviruses
85 East African cassava mosaic virus and French bean severe leaf curl virus, respectively. Whereas two ca.
86 7.6 kb sequences from the SRA SRX4959752 corresponding to RNA-Seq of *Welwitschia* mature leaves
87 from China [23] obtained best significant hits (E-value = 6e-125 and 5e-107, identity = 29.19% and
88 36.48%) with the polyprotein encoded by the unclassified *Caulimoviridae*-like virus *Pinus nigra* virus 1.
89 The tentative virus contigs were extended and curated by iterative mapping of the corresponding libraries
90 raw reads, using a method described by [24]. Essentially, this strategy employs BLAST/nhmmer to
91 extract a subset of reads related to the query contig, use the retrieved reads to extend it and then repeat the
92 process iteratively using as query the extended sequence. Importantly, the sequences extended and
93 polished presented overlapping regions and were subsequently reassembled using Geneious v8.1.9
94 (Biomatters Ltd.) alignment tool (cost matrix: >93% similarity 5.0/-9.026) into four circular single 2,913

nt (WMAgVA), 2,912 nt (WMAgVB), 7,872 nt (WMV1) and 7,648 nt long (WMV2) sequences. The circular nature of the virus genomes were evidenced by identifying overlapping reads simultaneously covering a continuum of virus reads supporting circularity at all genome positions. Bowtie2 <http://bowtie-bio.sourceforge.net/bowtie2/index.shtml> was then employed for mean coverage estimation and reads per million (RPMs) calculations. These resulting DNA virus sequences were supported by a total of 2,886 (WMV1), 13,119 (WMV2), 585 (WMAgVA), and 2,241 reads (WMAgVB) (mean coverage = 56.7X, 250.1X, 50.8X and 191.5X, respectively Supplementary Table 1). To further advance in the characterization of these sequences virus ORFs were predicted with ORFfinder (<https://www.ncbi.nlm.nih.gov/orffinder/>), domains presence and architecture of translated gene products was determined by InterPro (<https://www.ebi.ac.uk/interpro/search/sequence-search>) and the NCBI Conserved domain database v3.19 (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). Further, HHPred and HHBlits as implemented in <https://toolkit.tuebingen.mpg.de/#/tools/> were used to complement annotation of divergent predicted proteins by hidden Markov models. In order to infer phylogenetic relationships, the aa sequences of the WMV1, WMV2, WMAgVA and WMAgVB selected proteins (RT-RNase H domain for caulimovirids and CP and Rep for the geminiviruses), as well as the WMAgVA and WMAgVB complete genomic sequences, were aligned with the corresponding sequences of caulimovirids and geminiviruses described in Supplementary Table 2, using MAFFT 7 <https://mafft.cbrc.jp/alignment/software/> (BLOSUM62 scoring matrix) using as best-fit algorithm G-INS-i (RT, CP and Rep) and E-INS-i (complete genomic sequences). The aligned sequences were subsequently used as input for FastTree 2.1.9 (<http://www.microbesonline.org/fasttree/>) ML phylogenetic trees computing local support values with the Shimodaira-Hasegawa test (SH) and 1,000 tree resamples or MegaX [25] where 1000 bootstrap replicates were used to construct the ML trees. The assembled sequences corresponding to the complete genome sequence of WMV1, WMV2, WMAgVA and WMAgVB were deposited in GenBank under accession numbers BK061147, BK061148, BK061149 and BK061150, respectively. The genome of WMV1 and WMV2 were determined to be 7,872 and 7,648 nt in size, respectively, and they have one single large open reading frame (ORF) coding a 2,264 and 2,111 amino acids (aa) polyprotein, respectively (Fig. 1A). The WMV1 encoded polyproteins have several conserved domains: movement protein (aa 63 to 177 coordinates), reverse transcriptase (aa 1,419 to 1,667), and RT-RNase H (aa 1,689 to 1,810); whereas a movement protein (aa 57-188), Peptidase A3 (aa 932 to 1,080), reverse transcriptase (aa 1,124 to 1,302), and RNase H (aa 1,399 to 1,559) conserved domains were identified in the WMV2 polyprotein. Similar domains were described in the identified *Caulimoviridae* virus and best hit Pinus Nigra virus 1 (PNV1) [26] and other caulimovirids [19]. The WMV1 and WMV2 polyprotein shared just a 28.3% aa identity and their polyproteins shared the highest identity (36% and 29%, respectively) with that one encoded by PNV1; whereas the WMV1 and WMV2 RT- RNase H domain, which is currently used for the demarcation of species in the family *Caulimoviridae* [27] shared a 48.2% aa identity between them, and a highest identity of 54.14% and 52.78%, respectively, with the similar domain of PNV1. Therefore, the substantially low genetic identity of WMV1 and WMV2 between them and with other *caulimoviridae* members suggests that these two viruses should be classified as members of two new species.

In order to determine the phylogenetic relationships of WMV1 and WMV2 with other caulimovirids to unravel its possible evolutionary history, maximum likelihood (ML) phylogenetic analyses were carried out using aa sequences of the RT-RNase H domain. ML phylogenetic trees were constructed using G-INS-i as the best-fit model. Phylogenetic analysis showed that WMV1 and WMV2 clustered together with PNVI (Fig. 1B). PNVI has been proposed to represent a new genus of *Caulimoviridae*-infecting gymnosperms [26]. Thus, WMV1 and WMV2 should belong, along with PNVI, to this recently proposed novel genus within the *Caulimoviridae* family, which could be named as *Gymnovirus*, because all these viruses have been identified in gymnosperms and are phylogenetically related, so they likely share a common evolutionary history consistent with infection of gymnosperm hosts, an understudied and neglected clade of spermatophytes in plant virology.

The WMaGVA and WMaGVB genomes are monopartite and were determined to be 2,913 and 2,912 nt in length, respectively. Their circular DNA contains a nonanucleotide sequence TATATTAAC, that indicates the probable origin of virion-strand and DNA replication, slightly different than the nonanucleotide sequence TAATATTAC, which is found in most of the geminiviruses described so far [28]. Five putative proteins were identified in the circular ssDNA genome of WMaGVA and WMaGVB. The inferred genome organization of these viruses has three virion-sense (V1, V2 and V3) and two complementary-sense (C1 and C1:C2) genes with two intergenic regions (LIR and SIR (Fig. 2A). Similar to Curtoviruses [21], the two novel viruses have three ORFs in the virion-sense strand with the same genomic architecture. Whereas the number and genomic location of the ORFs in the complementary-sense of WMaGVA and WMaGVB is similar to that described for Becurtoviruses and Mastreviruses [21]. The genes on sense and complementary strands are separated by two intergenic regions, a relatively short one (SIR) that separates the genes V1 and C1:C2, and a long one (LIR) that separates the genes C1 and V3 (Fig. 2A). A LIR and a SIR is present in several genera belonging to the *Geminiviridae* family [21, 28].

A BLASTp search analysis of the predicted proteins of WMaGVA and WMaGVB showed that two of their five encoded proteins, encoded by V2 and V3 genes, had no significant sequence similarity to any others sequences in the database (Table 1). However, one transmembrane domain was identified in the WMaGVA and WMaGVB V2 ORF encoded protein; transmembrane domains were also predicted in the putative movement protein (MP) of several geminiviruses [29]. Furthermore, the ORF V2 is immediately upstream of the capsid protein (CP) (Fig. 2A), where the putative movement proteins are usually located in monopartite members of *Geminiviridae* family [28]; thus, V2 gene likely encodes a MP. No conserved motif or domains were identified in the product encoded by the V3 gene. The WMaGVA and WMaGVB ORF V1 (Fig.1) is similar to the CP encoded by mastreviruses. The replication-associated protein A (RepA) is encoded by the C1 gene; while the replicase protein (Rep) is expressed from C1 and C2 genes potentially by transcript splicing and contains the conserved motifs described for this protein in other geminiviruses [30]. A similar splicing of C1:C2 was also described for becurtoviruses, capulaviruses, grabloviruses and mastreviruses [21]. Unlike the V1 ORF encoded protein, the translation product of ORF C1:C2 is similar to the Rep encoded by capulaviruses. In summary, the V1 gene encoded protein was identified as the putative CP, while the C1:C2 encoded protein was identified as the putative Rep protein, while V1 likely encodes the putative movement protein and C1 encodes the RepA protein.

Genome-wide pairwise comparisons using the Sequence Demarcation Tool (SDT v1.2) [31] indicated that WMaGVA and WMaGVB share 77.2% identity between each other, and these viruses shared a maximum of 60.5% identity with all of other known geminiviruses (Supplementary Fig. 1A). In pairwise comparisons, the WMaGVA and WMaGVB CP aa sequences shared 84.32% identity between each other, and a highest identity (30.71%) with the mastrevirus sugarcane white streak virus (SWSV) (Supplementary Fig. 1B); whereas WMaGVA and WMaGVB Rep protein sequences shared 87.16% identity between each other, and a highest identity (47.3%) with the capulavirus alfalfa leaf curl virus (Supplementary Fig. 1C). These results showed that these viruses are distantly related to other geminiviruses. Therefore, the deeply low genetic identity of WMaGVA and WMaGVB with other geminiviruses suggests that these two viruses should be classified as members of a new species that could belong to a novel genus within family *Geminiviridae*, which could be named as *Welwitschiavirus*. Phylogenetic analysis of the WMaGVA and WMaGVB sequences together with the full-genome (or DNA-A sequence) of other known geminiviruses, using E-INS-i as the best-fit model, indicated that these two viruses formed a distinct clade that clustered with a group formed by the capulaviruses members as well the unassigned geminiviruses tomato apical leaf curl virus and tomato associated geminivirus 1 (Fig.2B). Phylogenetic analyses carried out using the aa sequences of the CP and Rep genes with G-INS-i as the best-fit model showed that WMaGVA and WMaGVB CP formed a distinct clade that clustered with the a group formed by mastreviruses (Fig. 2C); whereas the Rep protein based tree showed that WMaGVA and WMaGVB Rep protein formed a distinct clade that clustered with a group formed by capulaviruses members as well the unassigned geminiviruses tomato apical leaf curl virus and tomato associated geminivirus 1 (Fig. 2D), which further supports the results obtained when both the CP and Rep amino acid sequences were analyzed using SDT. Phylogenetic analysis placed WMaGVA and, WMaGVB in a unique taxon within geminiviruses; therefore it could be concluded that this virus represents an evolutionary distinct geminivirus 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 WMaGVA and WMaGVB genomes. Several geminiviruses with chimeric genomes architecture and phylogeny, such as WMaGVA and WMaGVB were identified [32-36]. Moreover, recently, it was shown that the CP sequences of geminiviruses are likely co-diverging with their insect vector [37]. The distinct phylogenetic cluster formed by WMaGVA and WMaGVB indicates that these viruses are likely transmitted by a vector that has not yet been described for geminiviruses, which in turn should involve an insect adapted to the peculiar Namib coastal desert, one of the driest and oldest desert regions of the world. It is apparent that the WMaGVA and WMaGVB are chimeric virus encoding Rep and CP with differential evolutionary histories because their CP and Rep phylogenies are not congruent; thus, it is likely that these viruses are a recombinant of mastre-like and capula-like ancestral viruses. However, when recombination analysis of full-length genomes were performed using RDP, GENECONV, MaxChi, BootScan, 3Seq, Chimera and SiScan statistics methods implemented in the RDP5 software [38], no recombination events were detected in the WMaGVA and WMaGVB genome sequences. This result could be explained by the fact that the non-recombinant descendants of WMaGVA and WMaGVB parents have not been described yet or no longer exist.

215 WMaGVA and WMaGVB display a geminivirus-like genomic size and organization, and encodes a CP
216 that is most closely related to that of mastreviruses, and a Rep protein that is most closely related to that
217 of capulaviruses. Thus, these viruses represent an evolutionary distinct lineage within the geminiviruses
218 and therefore they are unique novel members that may represent a new genus within the *Geminiviridae*
219 family.

220 Given the difficulties associated to identifying a proper host based only on metagenomics analyses of
221 NGS data, we extended our search using as query the genomic sequences of the characterized viruses
222 from the initial two libraries to the assemblies generated for the additional 28 experiments. Interestingly
223 we found blastn hits ranging from 96-100% identity in 20% of the assessed libraries for WMV1 and
224 WMV2, and blastn hits ranging from 93-100% identity in 90% of the libraries for WMaGVA and
225 WMaGVB. We confirmed these detections by assessing the raw data by read mapping using Bowtie2
226 with standard parameters against the consensus genomes (Supplementary Table 1) which was consistent
227 with the detection of these viruses in multiple independent samples from independent studies obtained
228 from wild and ex situ plants from the Namib Desert, China and USA, from diverse tissues, organs and
229 developmental stages. These results provide further robust evidence on the association of these viruses
230 with *Welwitschia* and suggest that the caulimo and geminiviruses reported here appear to be prevalent in
231 this host.

232 In summary, the analysis of public SRA data is a valuable source to detect potential novel DNA plant
233 viruses. Using this approach, we explored a glimpse into the DNA virome of the fossil plant *Welwitschia*
234 *mirabilis* which resulted in the identification and reconstruction of the complete genome of two novel
235 geminiviruses and two novel caulimovirids associated with *Welwitschia mirabilis*. This is the first report
236 of caulimo and geminiviruses infecting this ancient gymnosperm plant worldwide. Future studies should
237 assess many intriguing aspects on the biology and ecology of these viruses linked with such a peculiar
238 host in a hostile and restricted habitat, their prevalence, their vectors, and some clues about their potential
239 impact on their host.

240

241 **Acknowledgments**

242 We would like to express a sincere gratitude to the generators of the underlying data used for this work.
243 By following open access practices and supporting accessible raw sequence data in public repositories
244 available to the research community, they have promoted the generation of new knowledge and ideas.

245

246 **Statements and Declarations**

247 **Funding**

248 The authors declare that no funds, grants, or other support were received during the preparation of this
249 manuscript.

250 **Competing Interests**

251 The authors have no relevant financial or non-financial interests to disclose.

252 **Author Contributions**

Humberto Debat and Nicolas Bejerman contributed to the study conception and design, data analysis. The manuscript was written by both authors, who commented and reviewed it. All authors read and approved the final manuscript.

Data Availability

The virus genome sequences are deposited in database Genbank under accession numbers BK061147, BK061148, BK061149 and BK061150.

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. Roskov Y, Ower G, Orrell T, Nicolson D, Bailly N, Kirk PM, et al. (2019) Species 2000 & ITIS Catalogue of Life. In: Annual Checklist, 2019.
2. Bombi P, Salvi D, Shuuya T, Vignoli L, Wassenaar T (2021) Climate change effects on desert ecosystems: A case study on the keystone species of the Namib Desert *Welwitschia mirabilis*. Plos One. 16(11):e0259767.
3. Jürgens N, Oncken I, Oldeland J, Gunter F, Rudolph B (2021) *Welwitschia*: phylogeography of a living fossil, diversified within a desert refuge. Sci. Rep. 11:2385.
4. Koh SH, Li H, Admiraal R, Jones MGK, Wylie SJ (2015) Catharanthus mosaic virus: a potyvirus from a gymnosperm, *Welwitschia mirabilis*. Virus Res. 203:41–46.
5. Roossinck MJ, Martin D, Roumagnac P (2015) Plant virus metagenomics: advances in virus discovery. Phytopathol. 105:716–727.
6. Susi H, Filloux D, Frilander MJ, Roumagnac P, Laine AL (2019) Diverse and variable virus communities in wild plant populations revealed by metagenomic tools. PeerJ 7:e6140.
7. Zhang YZ, Chen YM, Wang W, Qin XC, Holmes EC (2019). Expanding the RNA virosphere by unbiased metagenomics. Annu Rev Virol 6:119–139.
8. Simmonds P, Adams MJ, Benko M, Breitbart M, Brister JR, et al. (2017) Consensus statement: virus taxonomy in the age of metagenomics. Nat. Rev. Microbiol. 15:161–168.
9. Dutilh E, Varsani A, Tong Y, Simmonds P, Sabanadzovic S, et al. (2021). Perspective on taxonomic classification of uncultivated viruses. Curr Opinion Virol 51:207-215.
10. Debat HJ, Bejerman N (2019) Novel bird's-foot trefoil RNA viruses provide insights into a clade of legume-associated enamoviruses and rhabdoviruses. Arch Virol. 164:1419-1426.
11. Edgar RC, Taylor J, Lin V, Altman T, Barbera P, et al. (2020) Petabase-scale sequence alignment catalyses viral Discovery. Nature. 602:142-147.
12. Lauber C, Seifert M, Bartenschlager R, Seitz S. (2019) Discovery of highly divergent lineages of plant-associated astro-like viruses sheds light on the emergence of potyviruses. Virus Res. 260:38-48.

13. Mifsud J, Gallagher R, Holmes E, Geoghegan J (2022) Transcriptome Mining Reveals a Spectrum of RNA Viruses in Primitive Plants. bioRxiv. <https://doi.org/10.1101/2022.02.07.479358>.
14. Nibert ML, Vong M, Fugate KK, Debat, HJ (2018) Evidence for contemporary plant mitoviruses. *Virology* 518:14–24.
15. Sidharthan K, Kalaivanan N, Baranwal V (2021) Discovery of putative novel viruses in the transcriptomes of endangered plant species native to India and China. *Gene*. 786:145626.
16. Alcalá-Briseño RI, Casarrubias-Castillo K, López-Ley D, Garrett KA, Silva-Rosales L (2020) Network analysis of the papaya orchard virome from two agroecological regions of Chiapas. *Mexico mSystems* 5(1):e00423-e1419.
17. Bejerman N, Trucco V, de Breuil S, Pardina PR, Lenardon S, Giolitti F (2017) Genome characterization of an Argentinean isolate of alfalfa leaf curl virus. *Arch Virol* 5:1–5.
18. Debat HJ, Ng TFF (2019) Complete genome sequence of a divergent strain of Tibetan frog hepatitis B virus associated with a concave-eared torrent frog (*Odorrana tormota*). *Arch Virol* 164:1727–1732.
19. Teycheney P-Y, Geering ADW, Dasgupta I, Hull R, Kreuze JF, et al (2020) ICTV Virus taxonomy profile: Caulimoviridae. *J General Virol* 10:1025–1026.
20. Harper G, Hull R, Lockhart B, Olszewski N (2002) Viral sequences integrated into plant genomes. *Annu. Rev. Phytopathol.*, 40, 119–136.
21. Zerbini FM, Briddon RW, Idris A, Martin DP, Moriones E, et al (2017) ICTV Virus Taxonomy Profile: Geminiviridae. *J Gen Virol* 98:131–133.
22. Kan S, Shen T, Ran J, Wang X (2021). Both Conifer II and Gnetales are characterized by a high frequency of ancient mitochondrial gene transfer to the nuclear genome. *BMC Biol.* 19:146.
23. Ran J-H, Shen T-T, Wang M-M, Wang X-Q. Phylogenomics resolves the deep phylogeny of seed plants and indicates partial convergent or homoplastic evolution between Gnetales and angiosperms. *Proc R Soc B.* 2018;285:20181012.
24. Singh KS, Beadle K, Troczka BJ, Field L, Davies E, Williamson M, Bass C (2018) Extension of partial gene transcripts by iterative mapping of RNA-Seq raw reads. *IEEE/ACM Trans Comput Biol Bioinform.* <https://doi.org/10.1109/tcbb.2018.2865309>.
25. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35:1547–1549.
26. Rastrojo A, Núñez A, Moreno DA, Alcamí A (2018) A new putative Caulimoviridae genus discovered through air metagenomics. *Microbiol. Resour. Announc.* 7:e00955-18.
27. Geering ADW, Hull R, (2012) Family Caulimoviridae. In: King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ. (eds) *Virus Taxonomy, Classification and Nomenclature of Viruses; Ninth Report of the International Committee on Taxonomy of Viruses*. London, UK, Elsevier Academic Press, pp 429–443.
28. Varsani A, Roumagnac P, Fuchs M, Navas-Castillo J, Moriones E, Idris A, Briddon RW, Rivera-Bustamante R, Murilo Zerbini F, Martin DP (2017) Capulavirus and Grablovirus: two new genera in the family Geminiviridae. *Arch Virol* 162:1819–1831.

29. 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.
30. Bernardo P, Golden M, Akram M, Naimuddin Nadarajan N, Fernandez E, Granier M, Rebelo AG, Peterschmitt M, Martin DP, Roumagnac P (2013) Identification and characterisation of a highly divergent geminivirus: evolutionary and taxonomic implications. *Virus Res* 177:35–45.
31. Muhire BM, Varsani A, Martin DP (2014) SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. *PLoS One* 9:e108277.
32. Claverie S, Bernardo P, Kraberger S, Hartnady P, Lefeuvre P, Lett JM, Galzi S, Filloux D, Harkins GW, Varsani A, Martin DP, Roumagnac P (2018) From spatial metagenomics to molecular characterization of plant viruses: a geminivirus case study. *Adv Virus Res* 101:55–83.
33. Fontenele RS, Lamas NS, Lacorte C, Lacerda ALM, Varsani A, Ribeiro SG (2017) A novel geminivirus identified in tomato and cleome plants sampled in Brazil. *Virus Res* 240:175–179.
34. Fontenele RS, Abreu RA, Lamas NS, Alves-Freitas DMT, Vidal AH, Poppiel RR, Melo FL, Lacorte C, Martin DP, Campos MA, Varsani A, Ribeiro SG (2018) Passion fruit chlorotic mottle virus: molecular characterization of a new divergent geminivirus in Brazil. *Viruses* 10:169.
35. Vaghi Medina CG, Teppa E, Bomancini VA, Flores CR, Marino-Buslje C, Lopez Lambertini PM (2017) Tomato apical leaf curl virus: A novel, monopartite geminivirus detected in tomatoes in Argentina. *Front Microbiol* 8:2665.
36. Zhang R, Wu X, Jiang X, Wu X, Luan X, Cheng X (2020) Molecular characterization of common bean curly stunt virus: a novel recombinant geminivirus in China. *Arch Virol* 165:257–260.
37. Lefeuvre P. et al. (2019) Evolution and Ecology of Plant Viruses. *Nature Rev Microbiol* 17:632–44.
38. Martin DP, Varsani A, Roumagnac P, Botha G, Maslamoney S, et al. (2020) RDP5: A computer program for analysing recombination in, and removing signals of recombination from, nucleotide sequence datasets. *Virus Evol.* 7:veaa087.

Fig. 1. (A) Genomic organization of Welwitschia mirabilis virus 1 (WMV1) and Welwitschia mirabilis virus 2 (WMV2). (B) Maximum likelihood phylogenetic tree analyzing the WMV1 and WMV2 and representative caulimovirids RT-RNase H domain amino acid sequences. The bar below each tree represents substitutions per site. Accession numbers of every virus used to construct phylogenetic trees are listed in Supplementary Table 2.

Fig. 2. (A) Genomic organization of Welwitschia mirabilis associated geminivirus A (WMaGVA) and Welwitschia mirabilis associated geminivirus B (WMaGVB). Maximum likelihood phylogenetic tree analyzing the WMaGVA and WMaGVB and representative geminiviruses complete genome nucleotide sequences (B), as well as CP (C) and Rep (D) amino acid sequences. The bar below each tree represents

372 substitutions per site. Accession numbers of every virus used to construct phylogenetic trees are listed in
373 Supplementary Table 2.

A

B

A

bioRxiv preprint doi: <https://doi.org/10.1101/2022.02.23.481439>; this version posted February 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

B

C

D

Table 1. Genomic organization of Welwitschia mirabilis associated geminivirus A (WMaGVA) and Welwitschia mirabilis associated geminivirus B (WMaGVB)

Virus	ORF/Region	Locus	Length nt/aa	Putative protein	Highest scoring virus- protein/E-value/query coverage% (Blast P)
WMaGVA	V1	520-1380	861/287	CP	SWSV/4e-08/65
	V2	350-643	294/98	possible MP	no hits
	V3	486-869	384/128	unknown	no hits
	C1	2829-1846	984/328	RepA	ALCV/9e-56/70
	C1:C2	2829-2221, 2029-1652	987/329	Rep	ALCV/6e-92/94
WMaGVB	V1	516-1376	861/287	CP	SWSV/3e-08/67
	V2	346-639	294/98	possible MP	no hits
	V3	482-877	396/132	unknown	no hits
	C1	2825-1845	981/327	RepA	ALCV/5e-57/70
	C1:C2	2825-2217, 2028-1651	987/329	Rep	ALCV/5e-89/94