

Article

Single Stranded DNA Viruses Associated with Capybara Faeces Sampled in Brazil

Rafaela S. Fontenele ^{1,2}, Cristiano Lacorte ², Natalia S. Lamas ², Kara Schmidlin ¹, Arvind Varsani ^{1,3,*}  and Simone G. Ribeiro ^{2,*} 

¹ The Biodesign Center for Fundamental and Applied Microbiomics, Center for Evolution and Medicine School of Life Sciences, Arizona State University, Tempe, AZ 85287, USA

² Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF 70770-017, Brazil

³ Structural Biology Research Unit, Department of Clinical Laboratory Sciences, University of Cape Town, Observatory, Cape Town 7925, South Africa

* Correspondence: arvind.varsani@asu.edu (A.V.); simone.ribeiro@embrapa.br (S.G.R.); Tel.: +1-480-727-2093 (A.V.); +55-61-3448-4666/3448-4703 (S.G.R.)

Received: 14 June 2019; Accepted: 31 July 2019; Published: 2 August 2019



Abstract: Capybaras (*Hydrochoerus hydrochaeris*), the world's largest rodents, are distributed throughout South America. These wild herbivores are commonly found near water bodies and are well adapted to rural and urban areas. There is limited information on the viruses circulating through capybaras. This study aimed to expand the knowledge on the viral diversity associated with capybaras by sampling their faeces. Using a viral metagenomics approach, we identified diverse single-stranded DNA viruses in the capybara faeces sampled in the Distrito Federal, Brazil. A total of 148 complete genomes of viruses in the *Microviridae* family were identified. In addition, 14 genomoviruses (family *Genomoviridae*), a novel cyclovirus (family *Circoviridae*), and a smacovirus (family *Smacoviridae*) were identified. Also, 37 diverse viruses that cannot be assigned to known families and more broadly referred to as unclassified circular replication associated protein encoding single-stranded (CRESS) DNA viruses were identified. This study provides a snapshot of the viral diversity associated with capybaras that may be infectious to these animals or associated with their microbiota or diet.

Keywords: *Hydrochoerus hydrochaeris*; CRESS DNA virus; *Circoviridae*; *Genomoviridae*; *Microviridae*; *Smacoviridae*

1. Introduction

Capybaras (*Hydrochoerus hydrochaeris*, Linnaeus 1766) are the world's largest rodent and have a wide distribution throughout South America. They are herbivores with a generalist diet, and their semi-aquatic habit requires the presence of water bodies such as lakes, rivers, and lagoons [1]. Capybaras are wild animals but also well adapted to human-modified landscapes [2,3]. Due to their high reproductive rates, capybaras can expand their population rapidly, especially in areas with few predators such as urban environments and agricultural farms [4,5]. The proximity of capybara habitats to humans and domestic animals in urban areas can facilitate the zoonotic spread of pathogens such as *Rickettsia rickettsii* [6].

A hand full of viruses have been identified associated with capybaras. Recently, a mimivirus (family *Mimiviridae*) was identified in faecal samples of 17 capybaras collected in the Midwest and Southeast regions of Brazil [7]. Vaccinia virus (VACV) (family *Poxviridae*) has been identified in faeces of capybaras from wild and urban areas of Brazil [8] and has also been detected in serum samples by serological assays [9]. A study with experimentally infected capybaras demonstrated that VACV

is able to replicate in this rodent species [10], supporting their role in the ecology of VACV [11]. Serological assays have shown the presence of Rabies virus (family *Rhabdovirus*), which is believed to be transmitted to the capybaras by blood-feeding bats [12]. Finally, following a diarrhoea outbreak from a group of capybaras in a city from São Paulo state, Brazil, a coronavirus (family *Coronaviridae*) was identified based on transmission electron microscopy [13].

With the advent of high throughput sequencing (HTS) technologies, the discovery and identification of known and novel viruses have increased significantly over the last decade. Currently, there are five recognised eukaryotic-infecting circular replication associated protein encoding single-stranded (CRESS) DNA virus families: *Bacillidnaviridae* [14], *Circoviridae* [15], *Geminiviridae* [16], *Genomoviridae* [17], and *Nanoviridae* [18]. Whereas, the prokaryotic-infecting CRESS DNA viral families are *Microviridae* [19] and *Inoviridae* [20], which infect bacteria and *Pleolipoviridae* [21] and *Smacoviridae* [22]—recently identified to likely infect archaea due to the presence of CRISPR spacers identified by bioinformatics analysis [23]—that infect archaea. In addition, there are numerous novel clusters of CRESS DNA viruses that remain to be taxonomically classified. The replication associated proteins (Reps) of the eukaryotic circular Rep encoding single-stranded CRESS DNA viruses are distantly related to prokaryotic infecting ones.

With the aim to identify small DNA viruses associated with the capybara, we used a viral metagenomics approach on two faecal samples. A total of 201 viral genomes were identified; 164 span four known viral families and 37 are part of the unclassified CRESS DNA virus group.

2. Materials and Methods

2.1. Sample Collection and Processing

Two capybara faecal samples were collected in Brasilia and Planaltina, Distrito Federal, Brazil, in 2016. The fresh faecal pellet samples were picked off the ground from a grass field where wild capybaras were feeding, and were placed into a 50 mL tube. The faecal pellets were homogenised in an SM buffer (0.1 M NaCl, 50 mM Tris/HCl-pH 7.4, and 10 mM MgSO₄) and subsequently centrifuged for 10 min at 4300 rpm. The supernatant was sequentially filtered through a 0.45 µm and 0.2 µm syringe filter and PEG-precipitated (15% *w/v*) overnight. The precipitated filtrate was centrifuged, and the pellet was resuspended in 1 mL of SM buffer. A measure of 200 µL of this suspension was used for viral DNA extraction using the Zymo viral purification kit (Zymo Research, Irvine, CA, USA). The extracted viral DNA was enriched for circular viral DNA using rolling circle amplification (RCA) with the Illustra TempliPhi amplification kit (GE Healthcare, Chicago, IL, USA).

2.2. High Throughput Sequencing and Data Analysis

The RCA products from the two faecal samples were individually sequenced on an Illumina HiSeq 2500 platform (2 × 100 paired-end library) at Macrogen Inc., Seoul, South Korea. The paired-end reads were de novo assembled using SPAdes v 3.12.0 [24], and the resulting contigs (>750 nts) were analysed by BLASTx [25] against a local RefSeq viral protein sequence database. Based on the sequence of the de novo assembled contigs that had similarity with viral sequences, abutting primers were designed to recover the full-length genome by PCR (Supplementary Table S1), with the exception of the viral full-length genome contigs (determined based on terminal redundancy of de novo assembled contigs) that belong to the family *Microviridae*. PCR was performed using HiFi HotStart DNA polymerase (KAPA Biosystems, Wilmington, MA, USA) following the manufacturer's thermal cycling condition recommendations. The amplified genome sequences were resolved in a 0.7% agarose gel, and the expected size amplicons were gel-excised and purified using the Quick-spin PCR Product Purification Kit (iNtRON Biotechnology, Seongnam-si, South Korea). Gel purified amplicons were cloned into a pJET1.2 cloning vector (ThermoFisher Scientific, Waltham, MA, USA), and transformed into *E. coli* XL blue competent cells. The recombinant plasmids in the transformants were purified with the DNA-spin Plasmid DNA Purification kit (iNtRON Biotechnology, Seongnam-si,

South Korea) and Sanger sequenced by primer walking at Macrogen Inc. (Seoul, South Korea). Sanger sequence contigs were assembled and analysed in Geneious 11.1.5 [26]. For the full genomes assembled for viruses belonging to the *Microviridae* family, the raw reads were aligned using BWA v0.7.12 [27] for coverage depth assessment. The raw read data is deposited in the SRA database (PRJNA521956). All genomes recovered in this study were deposited in the GenBank database (Accession numbers MK496679-MK496826; MK483072-MK483085; MK570163-MK570200; MK947371; Supplementary Table S1).

2.3. Sequence Similarity Network Analyses

A dataset (Rep_all) of the Rep proteins of the alphasatellites, circoviruses, geminiviruses, genomoviruses, nanoviruses, and smacoviruses was assembled with sequences available in GenBank. A dataset of the major capsid protein (MCP) of the microvirus sequences available in GenBank was assembled (MCP_all). Both Rep_all and MCP_all were then separately clustered with a 0.9 sequence identity cut-off using CD-HIT [28]. For the Rep analysis, a representative from each cluster and all the Reps encoded by the viruses (except microviruses) identified in this study were assembled into a dataset (Rep_90_Cap). Similarly, for the MCP, a representative from the cluster was assembled into a dataset with all the MCP sequences (MCP_90_Cap) of the microviruses identified in this study. A sequence similarity network analysis using EST-EFI [29,30] with a minimum similarity score of 60 was constructed for the Rep_90_Cap dataset and with a minimum similarity score of 200 for the MCP_90_Cap dataset. The resulting sequence similarity networks were visualised in Cytoscape V3.7.1 [31] using the organic layout.

2.4. Sequence Analysis

All genome-wide and protein pairwise identities were determined using SDT v1.2 [32]. BLASTp [25] analysis of the MCP of the microviruses was undertaken to determine the closest related protein sequences using the complete MCP (MCP_all) dataset.

2.5. Circoviruses, Genomoviruses, and Smacoviruses

The Rep amino acid sequences within the circovirus, genomovirus, and smacovirus clusters were separately aligned by MUSCLE [33]. The alignment was then used to infer a Maximum Likelihood (ML) phylogenetic tree using PhyML 3.0 [34] with rtREV+G+I (for circoviruses), rtREV+G+I+F (for genomoviruses), and rtREV+G+I+F (for smacoviruses) amino acid substitution models, inferred as best fit models using ProTest [35], and an approximate likelihood ratio test (aLRT) was used for branch support. The cyclovirus Rep ML phylogenetic tree was rooted with the representative sequences of the unclassified CRESS DNA group (Cluster 1; Figure 4). The smacovirus ML phylogenetic tree has been rooted with Rep sequences of nanoviruses. Branch support <0.8 aLRT support was collapsed using TreeGraph2 [36].

2.6. Unclassified CRESS DNA Viruses

Based on the network analysis, clusters with ≥ 4 sequences that contained the capybara CRESS DNA virus Rep sequences were phylogenetically analysed. Sequences in each cluster were aligned by MUSCLE [33]. The alignments were used to infer the ML phylogenetic tree for each cluster using PhyML 3.0 [34] with the amino acid substitution model (Clusters 1: rtREV+G+I; 2: rtREV+G+I; 3: rtREV+G+I+F; 4: rtREV+G+I+F; 5: WAG+G+I; 6: rtrev+G; 7: WAG+G; 8: rtRev+G; 9: WAG+I; 10: rtREV+G+I; and 11: WAG+I) determined as a best fit model using ProTest [35], and an approximate likelihood ratio test (aLRT) was used for branch support. All ML phylogenetic trees were midpoint rooted, and branches with <0.8 aLRT support were collapsed using TreeGraph2 [36].

2.7. Microviruses

Based on the network analysis, the MCP of the microviruses in the clusters with ≥ 5 sequences containing the MCP from the capybara associated microviruses were analysed. MCP sequences of each cluster were aligned using MUSCLE [33]. The resulting alignments were used to infer an ML phylogenetic tree using PhyML 3.0 [34] with amino acid substitution models (Clusters 1: LG+I+G+F; 2: rtRev+G+I+F; and 3: rtRev+G) determined as a best fit model using ProTest [35], with an approximate likelihood ratio test (aLRT) for branch support. All ML trees were midpoint rooted and branches with < 0.8 aLRT support were collapsed using TreeGraph2 [36].

3. Results and Discussion

Two faecal samples from capybaras were collected in urban areas from Distrito Federal, Brazil and using HTS technologies, 201 CRESS DNA viruses were identified. Of these, 164 span four known viral families (*Genomoviridae*, *Circoviridae*, *Smacoviridae*, and *Microviridae*), and 37 are part of the unclassified CRESS DNA virus group (Figure 1; Supplementary Table S1). The group of CRESS DNA viruses has increased drastically in recent years due to the broad application of metagenomics approaches. However, most of the CRESS DNA viruses' genomes remain unclassified due to their high diversity and lack of host information.

3.1. Classified CRESS DNA Viruses

3.1.1. Genomoviruses

The family *Genomoviridae* was recently established and currently consists of nine genera [17,37]. The first isolated member of this family was the *Sclerotinia sclerotiorum* hypovirulence associated DNA virus 1 (SsHADV-1) [38]. Thus far, SsHADV-1 is the only genomovirus identified that is associated with a host, the fungus *Sclerotinia sclerotiorum*, which makes it the only fungal infecting ssDNA virus ever described [38]. The established genera within the *Genomoviridae* family are *Gemycircularvirus*, *Gemyduguivirus*, *Gemygorvirus*, *Gemykibivirus*, *Gemykolovirus*, *Gemykrogvirus*, *Gemykronzavirus*, *Gemytondvirus*, and *Gemyvongvirus* [37]. The genera demarcation has been established based on the Rep phylogenetic analysis. Within these genera, the current species demarcation threshold is 78% genome-wide pairwise identity. Genomoviruses have been isolated from a variety of environments such as fungi, plants, sediments, sewage/wastewater, insects, birds, and mammals faeces [37].

The capybara-associated genomoviruses ($n = 14$) identified in this study were all recovered from one of the capybara faecal samples (Cap1). These have been named capybara genomovirus 1–13 (CapGV1–13), with CapGV 2 having two variants, cap1_52 and cap1_64, sharing 98% genome-wide identity. These 14 genomes have classical features of genomoviruses with a conserved nonanucleotide motif at the origin of replication, and they encode a capsid protein (*cp*) in the virion sense and a replication-associated protein (*rep*) in the complementary sense (Supplementary Table S1; Supplementary Figure S1). The Rep encoded by the capybara genomoviruses all contain the rolling circle replication (RCR) endonuclease and superfamily 3 (SF3) helicase motifs conserved within genomoviruses Reps (Supplementary Table S2) [37]. The mapping of raw reads from each sample (Cap1 and Cap3) to the full-length genome sequences of the capybara-associated genomoviruses (Figure 1) confirmed that these genomoviruses were only present in sample Cap1, which is from where they were originally isolated.

A genome-wide pairwise identity of the CapGVs with representative sequences from all known species of genomoviruses was undertaken to identify them at the species level (Supplementary Data 1). CapGV1–CapGV6, CapGV8, CapGV10, CapGV12, and CapGV13 all represent novel genomovirus species with a genome-wide identity ranging from 60% to 75% with other classified genomoviruses. Phylogenetic analysis of the Rep sequences of the genomoviruses identified in this study reveals that

they can be classified within four established genera, *Gemycircularvirus* ($n = 3$), *Gemydunguivirius* ($n = 4$), *Gemykibivirius* ($n = 5$), *Gemykronzavirius* ($n = 1$), and one likely to an unclassified genus (Figure 2).

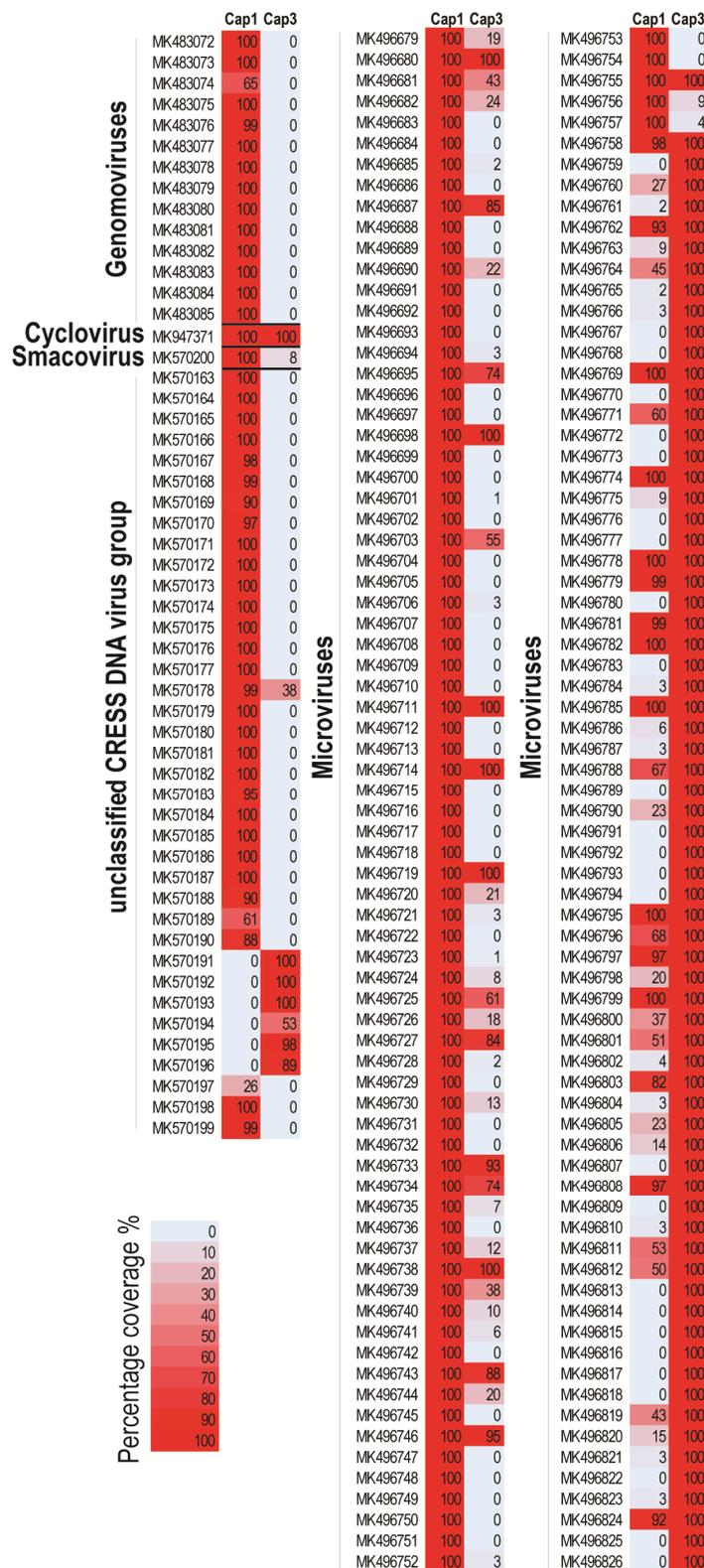


Figure 1. Percentage coverage of raw reads mapped from samples Cap1 and Cap3 against all genomes recovered in this study.

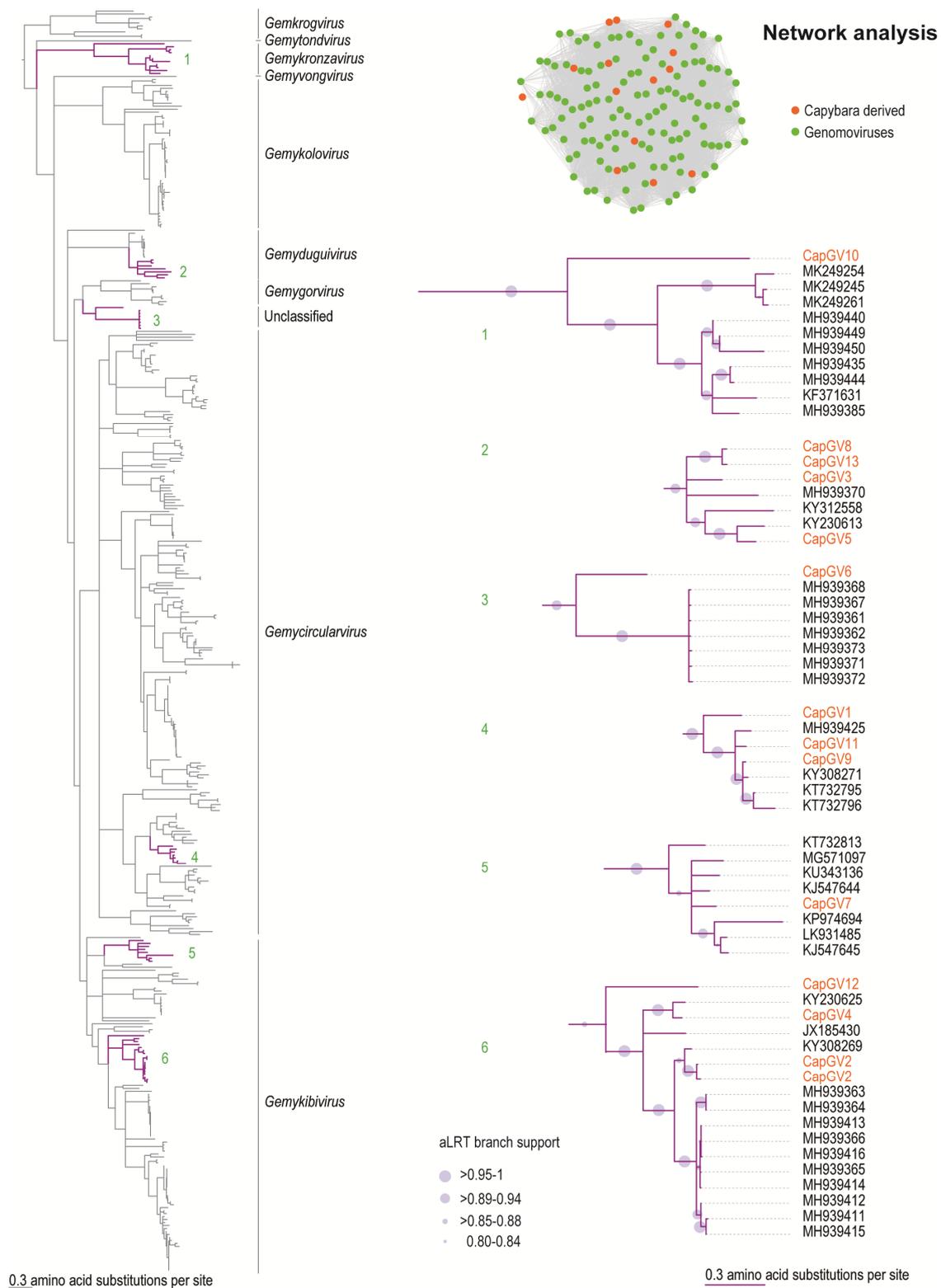


Figure 2. Sequence similarity network analysis and Maximum Likelihood phylogenetic tree of the Rep amino acid sequences of the viruses belonging to the nine *Genomoviridae* genera. The tree is rooted with Rep sequences of geminiviruses. Branches with aLRT support <0.8 were collapsed. The Reps of genomoviruses identified in this study are highlighted in orange. Clades containing the capybara associated genomoviruses CapGV1–13 (highlighted in purple) are expanded.

CapGV7 is a strain of sewage-associated gemycircularvirus-10a (KJ547644) [39], sharing 91% genome pairwise identity. CapGV9 is a new strain of the thrips associated genomovirus 2 (KY308271) [40], which shares 90% genome-wide identity. CapGV11 is a new strain of the plant associated genomovirus 12 (MH939425), sharing 84% genome-wide pairwise identity.

The genomovirus CapGV10 (MK483082) groups within the genus *Gemykronzavirus* (Figure 2) and has a nonanucleotide motif “TAAGATTCT.” The highest Rep and CP amino acid sequence identity is 51.3% and 34.9%, respectively, with a plant associated gemycircularvirus (MH939440) and CapGV2 (MK483073). CapGV3 (MK483075), CapGV5 (MK483077), CapGV8 (MK483080), and CapGV13 (MK483085) group with members of the genus *Gemydunguivirus* (Figure 2) and have variable nonanucleotide motifs “TAAKATTMT.” The CP and Rep amino acid sequences of CapGV3, CapGV13, and CapGV8 share 55.1%–65.3% and 85%–96% identity amongst themselves. The CP of CapGV5 shares 54.1% with that of the dragonfly associated circular virus 3 (JX185428) [41], whereas its Rep shares 86.6% identity with that of the Bemisia associated genomovirus AdDF (KY230613) [42], which was identified in Brazil.

The genomovirus CapGV6 (MK483078) has the nonanucleotide motif “TAATGTTAT” and does not group within any of the current nine established genera (Figure 2). Its Rep shares 63% amino acid identity with the plant-associated genomovirus 2 (MH939415), whereas its CP shares 72.3% sequence identity with the Pacific flying fox faeces associated gemycircularvirus 3 (KT732794) [43].

CapGV1 (MK483072), CapGV9 (MK483081), and CapGV11 (MK483083) group within the genus *Gemycircularvirus* (Figure 2) and have the nonanucleotide motif “TAATRITAT.” The Rep and CP of CapGV11 share 94.7% and 78.9% amino acid sequence identity, respectively, with the plant-associated genomovirus 12 isolate LT 2029 (MH939425). The Rep and CP of CapGV9 share 97.9% and 85.9% amino acid sequence identity, respectively, with the thrips associated genomovirus 2 (KY308271) [40]. The CP of CapGV1 shares 53.3% identity with the CP of the Pacific flying fox faeces associated gemycircularvirus 5 (KT732797) [43], whereas its Rep shares 78.4% sequence identity to that of CapGV11 (MK483083) from this study.

The genomoviruses that group within the *Gemykibivirus* genus (Figure 2) are the CapGV12 (MK483084), CapGV4 (MK483076), CapGV2 (MK483073–MK483074), and CapGV7 (MK483079), which have a nonanucleotide motif of “TAATRITAT.” The Reps of CapGV2 share 92–93% identity with that of the thrips associated genomovirus 3 (KY308269) [40], and their CPs share 53% identity with that of the Pacific flying fox faeces associated gemycircularvirus 10 (KT732805) [43]. The Rep of CapGV4 shares 93% identity with that of the Bemisia associated genomovirus (KY230625) [42], and its CP shares 69% identity with the plant-associated genomoviruses (MH939366–MH939414). The Rep and CPs of CapGV7 share 93% and 96% amino acid sequence identity with that of the sewage-associated gemycircularvirus 10a (KJ547644) [39]. The CP of CapGV12 shares 72.8% identity with the finch associated genomovirus 2 isolate E50P_A (MK249296) [44], and its Rep shares 68.3% identity with the finch associated genomovirus 2 isolate E50N_A (MK249293) [44].

3.1.2. Circoviruses

The family *Circoviridae* is composed of animal-infecting ssDNA viruses with genomes of ~1.8–2.4 kb. Currently, the family is divided into two genera: *Circovirus* and *Cyclovirus*. Cycloviruses have been associated with both vertebrates and invertebrates, whereas circoviruses seem to be restricted to vertebrates [45].

The Rep of the capybara associated cyclovirus (CaCyV; MK947371) identified from the capybara faeces samples clusters with those of the family *Circoviridae* (Figure 3A). The circular genome of CaCyV has 1897 nucleotides, and it encodes a replication-associated protein on the complementary sense and a capsid protein in the virion sense (Figure 3D) with the conserved nonanucleotide motif “TAGTATTAC.” Further analysis of the Rep encoded by CaCyV showed they contain the conserved RCR endonuclease and SF3 helicase motifs presented in the Rep of members of the family *Circoviridae* (Supplementary

Table S2) [45]. Mapping of raw reads from each sample (Cap1 and Cap3) against the full-length genome of the CaCyV (Figure 1) revealed that the cyclovirus is common to both samples.

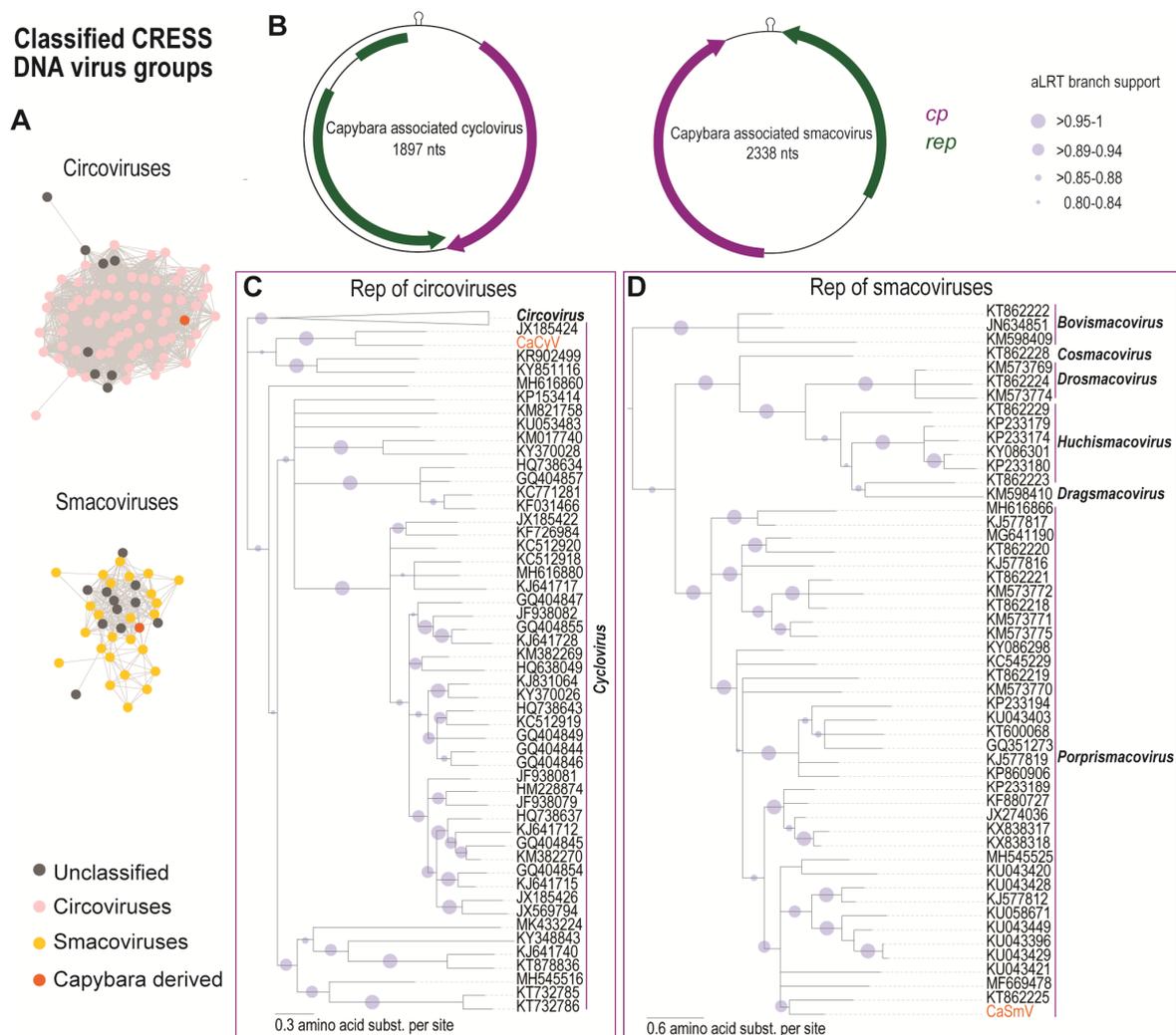


Figure 3. (A) Sequence similarity network analysis of the Rep amino acid sequences of viruses in the families *Circoviridae* and *Smacoviridae*. Orange dots represent the Rep of the viruses identified in this study. (B) Genome organisation of the capybara associated cyclovirus and capybara associated smacovirus. (C) Maximum Likelihood (ML) phylogenetic tree of the Rep amino acid sequences of the capybara associated cyclovirus with representative Rep sequences from viruses in the family *Circoviridae*. The cyclovirus Rep ML phylogenetic tree is rooted with representative sequences of unclassified CRESS DNA group (Cluster 1; Figure 4) and branches with aLRT support <0.8 have been collapsed. (D) ML phylogenetic tree of the Rep amino acid sequences of the capybara associated smacovirus with representative Rep sequences from viruses in the family *Smacoviridae*. The ML tree has been rooted with Reps of nanoviruses, and branches with aLRT support <0.8 were collapsed.

The Rep amino acid sequence phylogenetic analysis reveals that the new sequence belongs to the genus *Cyclovirus* (Figure 3B). CaCyV encodes a spliced Rep and represents a new species within the family sharing the highest genome-wide sequence pairwise identity of 62% (species demarcation for circoviruses is 80% genome pairwise identity [45]) with the dragonfly associated cyclovirus 5 (JX185426) [41] (Supplementary Data 2). The Rep of CaCyV shares 62.5% identity with that of the dragonfly associated cyclovirus 3 (JX185424) [41], and its CP shares 33.6% identity with that of the bat-associated cyclovirus 3 (JF938081) [46] (Supplementary Data 2). Several other cycloviruses have

also been identified in animal faeces [43,46–49], suggesting that these viruses may be associated with their diet.

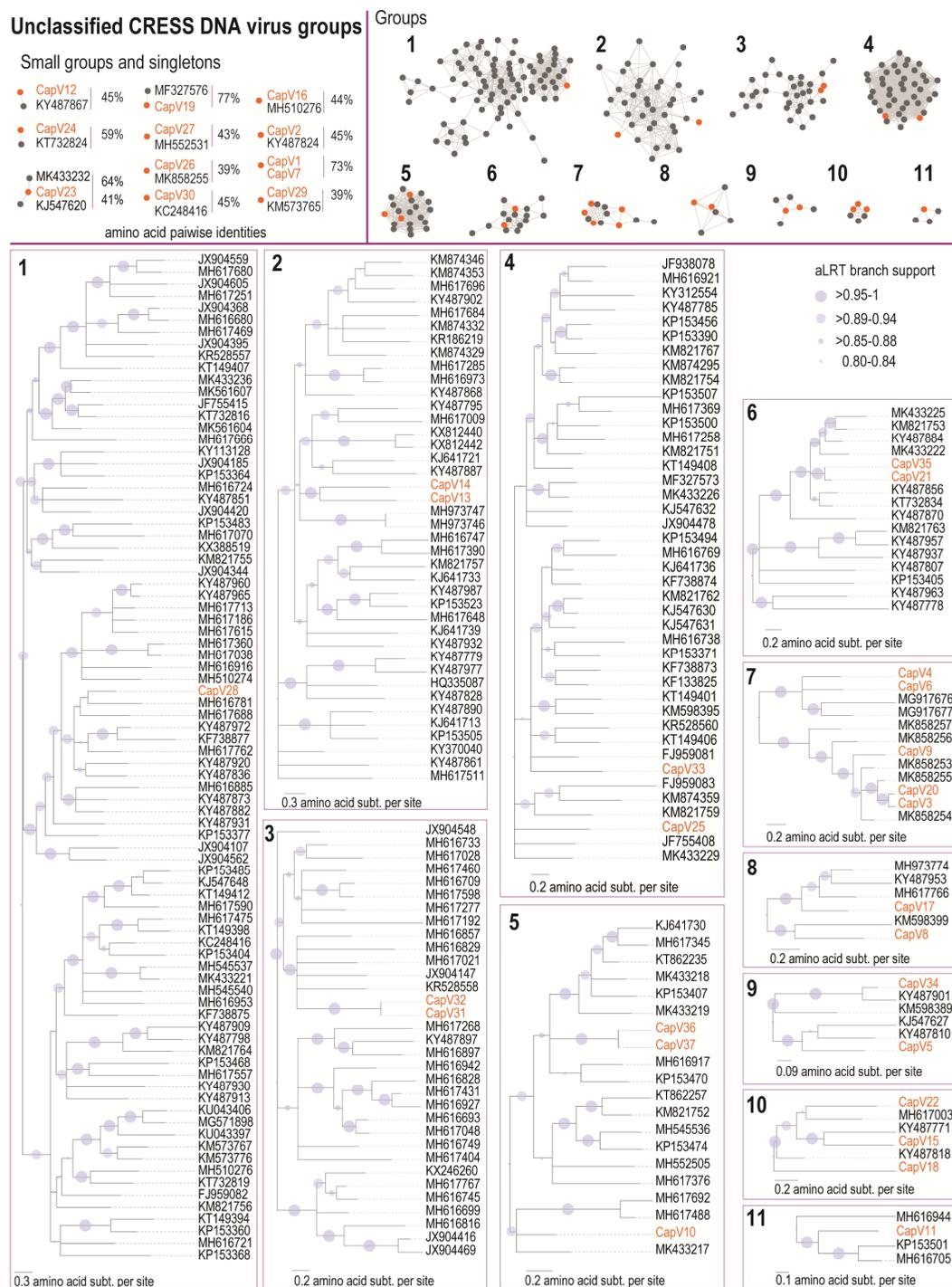


Figure 4. Sequence similarity network analysis and maximum likelihood phylogenetic trees of the Rep amino acid sequences of the unclassified CRESS viruses associated with the capybara. Orange dots represent the Rep of the unclassified CRESS DNA viruses identified in this study. The capybara viruses 1–37 are highlighted in orange, and branches with aLRT support <0.8 have been collapsed.

3.1.3. Smacoviruses

Viruses in the family *Smacoviridae* have circular ssDNA genomes of ~2300–2500 nucleotides. Smacoviruses encode a Rep and CP that are bidirectionally transcribed with a conserved nonanucleotide

at the origin of replication. This recently established family is classified into six genera *Bovismacovirus*, *Cosmacovirus*, *Dragsmacovirus*, *Drosmacovirus*, *Huchismacovirus*, and *Porprismacovirus* [22]. Smacoviruses have been primarily identified through the metagenomics analysis of diverse animal faecal samples. To date, these viruses have not been cultured nor has a conclusive host association been identified. A recent study provides some evidence that smacoviruses may infect archaea based on CRISPR spacers with smacovirus-like sequences identified in archaea [23].

In this study, one smacovirus (Capybara associated smacovirus, CaSmV; MK570200) was identified in the capybara faeces with a 2338 nts genome and a conserved nonanucleotide “TAGTGTTAC.” The genome encodes a Rep and a CP with two intergenic regions (Figure 3D). Mapping of raw reads from each sample (Cap1 and Cap3) to the full-length genome of the CaSmV (Figure 1) revealed that it is only present in sample Cap1, where it was initially recovered from. The Rep contains the RCR endonuclease and the SF3 helicase domain conserved within the Reps of smacoviruses [50] (Supplementary Table S2). Rep-based phylogenetic analysis reveals that CaSmV can be classified in the genus *Porprismacovirus* (Figure 3C). The CaSmV genome sequence is most closely related to the *Macaca mulatta* faeces associated virus 7 (KU043421) [51], sharing 62.7% genome pairwise identity (Supplementary Data 3). Based on the species demarcation threshold of 77% genome pairwise identity for smacoviruses [22], CaSmV is representative of a new species. The Rep of CaSmV shares 61% amino acid identity with that of the turkey associated porprismacovirus 1 (KF880727) [52], whereas the CP shares 42.3% identity with the chicken associated smacovirus (MG846353) [53].

3.2. Unclassified CRESS DNA Viruses

There are numerous reports of diverse CRESS DNA viruses that cannot be classified into currently established viral families. They all encode at least a Rep and a CP with different genome size and organisation (Supplementary Table S1; Supplementary Figure S2). From the two capybara faecal samples, 37 genomes were identified that cannot be classified into known families, and thus fall into the unclassified CRESS DNA virus group. These have been designated by the names capybara virus 1–37 (CapV 1–37). The CapVs range in size of 1525 to 3035 nts (Supplementary Figure S2). Out of the 37 genomes, 31 were isolated from sample Cap1, and 6 from Cap3. Mapping of raw reads from each sample (Cap1 and Cap3) to the full-length genome of the capybara unclassified CRESS DNA viruses (Figure 1) revealed that only the CapV16 genomic sequence (MK570178) that was isolated from Cap1 is common in both samples.

The Reps encoded by the CapVs contain the RCR endonuclease and SF3 helicase motifs that are conserved among the Reps of CRESS DNA viruses, with the exception of CapV24, which is missing the N-terminal RCR motifs I and motif II (Supplementary Table S2). Based on the sequence similarity network analysis of the Rep amino acid sequences, 24 CapVs cluster with other Reps forming 11 groups with ≥ 4 sequences (Figure 4). The remaining 13 sequences cluster in smaller groups or are singletons. The Rep of CapV28 clustered in group 1 shares 58.9% amino acid sequence identity with that of the CRESS DNA virus from rainbow trout tissue (MH617762) [54] (Supplementary Data 4). In group 2, the Reps of CapV13 and CapV14 share 47% and 44% amino acid sequence, respectively, with the Rep of a CRESS DNA virus sequence from wastewater (KY487868) [55] (Supplementary Data 4). Group 3 contains Reps of CapV31 and CapV32 that share 99.6% Rep amino acid identity among themselves and 48.5% with that of a Fiddler Crab associated circular virus (KR528558) [56] (Supplementary Data 4). In group 4, the Reps of CapV33 and CapV25 share 55–58% amino acid identity with that of the sewage-associated circular DNA virus 7 (KJ547631) [39] (Supplementary Data 4). In group 5, CapV10, CapV36, and CapV37 all have unidirectionally organised open reading frames. The Reps of CapV36 and CapV37 share 99.6% amino acid identity amongst themselves and 62.3% identity with that of the Lake Sarah-associated circular virus 34 (KP153470) [57] (Supplementary Data 4). The Rep of CapV10 shares 65% amino acid identity with the blackfly DNA virus 3 (MK433217) [58] (Supplementary Data 4). CapV21 and CapV35 are part of group 6, and their Reps share 92.6% amino acid identity amongst themselves and 67–72.6% with that of the Pacific flying fox faeces associated circular DNA virus 15

(KT732834) [43] (Supplementary Data 4). Group 7 contains five CapVs (CapV3, CapV4, CapV6, CapV9, and CapV20), all of which contain a putative spliced Rep and bidirectionally organised ORFs. The Reps of CapV3 and CapV20 share 89% amino acid identity. The Rep of CapV9 shares 69.6% amino acid identity with that of a tortoise associated circular virus (MK858253), and the Reps of CapV6 and CapV4 share 64% identity among themselves (Supplementary Data 4).

In group 8, the Rep of CapV17 shares 59.7% identity with that of the *Apis mellifera* virus 5 (MH973774) [59], while the Rep of CapV8 shares 48.6% identity with that of the Odonata associated circular virus 2 (KM598399) [60] (Supplementary Data 4). The Reps of CapV5 and CapV34 cluster in group 9 and the Rep of CapV5 shares 55.9% amino acid identity with that of a Rep from a CRESS DNA virus identified in wastewater (KY487810) [55], whereas that of CapV34 shares 76% identity with that of another CRESS DNA virus sequence from wastewater (KY487901) [55] (Supplementary Data 4). In group 10, the Rep of CapV15 shares 51.5% amino acid identity with the Rep of a CRESS DNA virus identified in wastewater (KY487771) [55] (Supplementary Data 4). The Reps of CapV18 and CapV22, both in group 10, share 48.8–52.7% identity with the Rep of a CRESS identified in wastewater (KY487818) [55] (Supplementary Data 4). Finally, in group 11, the Rep of CapV11 shares 51% identity with the Lake Sarah-associated circular virus 45 (KP153501) [57] (Supplementary Data 4). The smaller CapVs groups and singletons and their closest Rep identity to other CRESS DNA virus Reps (39–77%) are shown and summarised in Figure 4.

3.3. Bacteriophages

Microviruses

The bacterial-infecting CRESS viral family *Microviridae* is divided in two subfamilies, *Gokushovirinae* and *Bullavirinae*. They have been identified in a variety of environments, including animal gut and faecal samples [61–63], insects [58,59], sediments [64], seawater [65,66], and freshwater [67]. The viruses in the subfamily *Bullavirinae* are known to infect *Escherichia coli* [68]. Some viruses in the subfamily *Gokushovirinae* are known to infect *Spiroplasma*, *Chlamydia*, and *Bdellovibrio* [69–71]. However, there is no host association for most of those identified through viral metagenomics.

The microviruses associated with capybaras likely infect their gut microbiota or are associated with their diet. Within the two capybara faecal samples from this study, 148 microvirus genomes were identified, and of these, 80 were identified in Cap1 and 68 in Cap3. Mapping of the raw reads from each sample (Cap1 and Cap3) to the full-length genome of the capybara microviruses reveals that 36 genomic sequences are common to both samples (Figure 1). The genomes of the capybara microviruses range from 4148 to 6887 nts in size, and most encode at least the major capsid protein (MCP; pfam PHA00363), minor capsid protein (pfam PHA00327), and replication initiation protein (pfam PHA00330) (Supplementary Figure S3).

The cluster analysis of the MCP, the most conserved protein of microviruses, reveals that the 148 microviruses identified in this study belong to the subfamily *Gokushovirinae*. The capybara microviruses are highly diverse with the majority ($n = 140$) of them forming groups ≤ 3 or as singletons (Figure 5A). Nonetheless, some sequences cluster with MCPs of microviruses identified in faecal samples (see Figure 5A). We identified three main clusters that have the capybara associated microviruses MCPs. Cluster 1 is composed of MCPs of microviruses identified in terrestrial vertebrates, with the majority of them identified in mammals. Cluster 2 and 3 MCPs are composed of those microviruses identified in faeces of mammals.

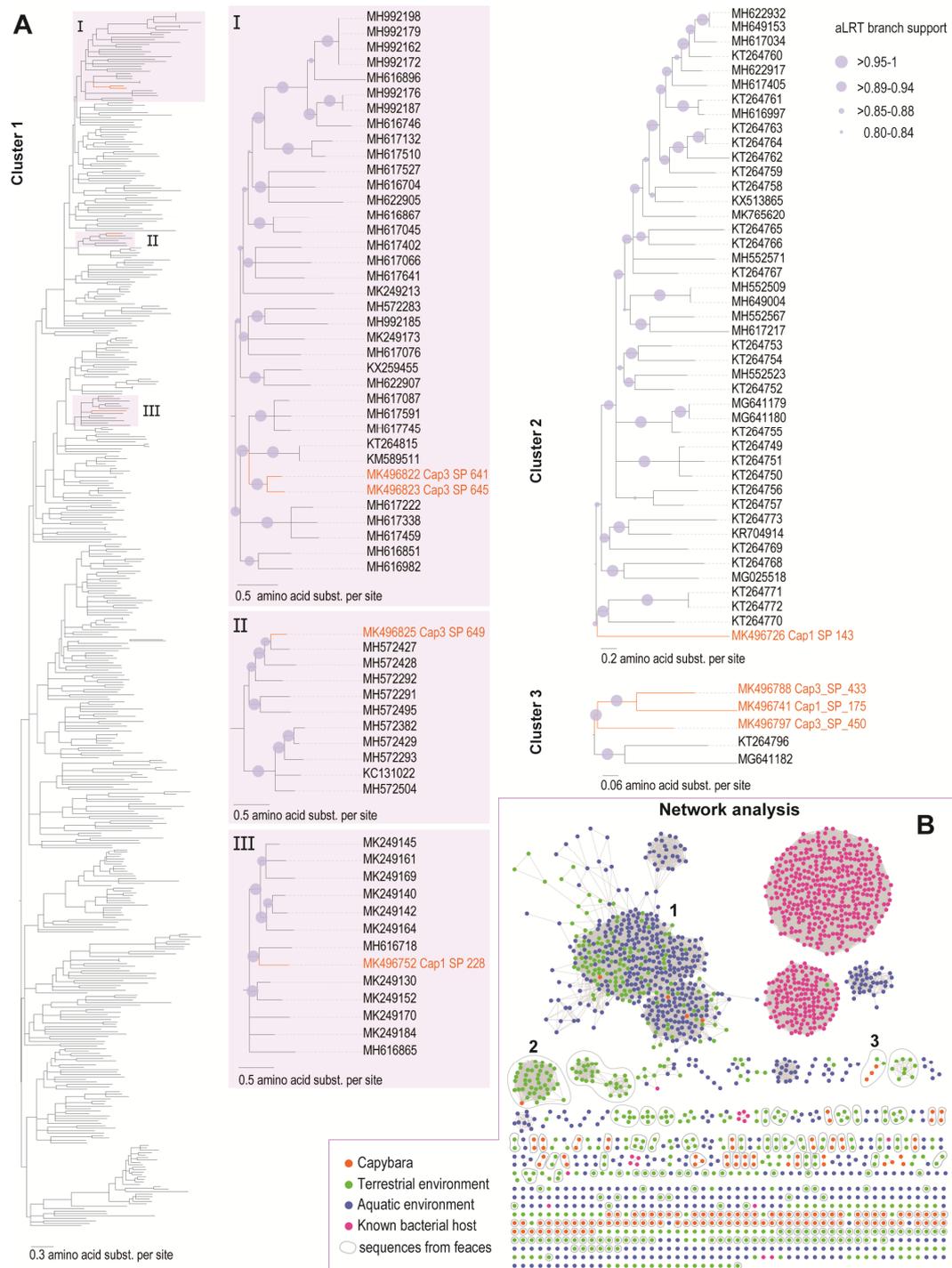


Figure 5. (A). Maximum Likelihood phylogenetic trees of the three main clusters (≥ 5 sequences) containing the capybara associated microviruses. The ML phylogenetic tree of cluster 1 has the branches in purple expanded (I, II, III). **(B).** Sequence similarity network of the major capsid protein (MCP) amino acid sequences of the capybara associated microviruses with a representative dataset of microviruses (MCP_all). Sequences are color-coded based on the type of environment they were recovered (terrestrial in green, aquatic in blue, and from known bacterial hosts in pink). The capybara samples are represented by orange dots, and all sequences derived from faeces are circled in grey.

The high diversity of the capybara associated microviruses was further confirmed with a pairwise amino acid comparison of their encoded MCP (Supplementary Data 5). The MCPs of the capybara associated microviruses share 16% to 98% amino acid sequence identity amongst themselves. From

the 148 capybara associated microviruses, the most closely related at the MCP level are the isolate Cap3_SP_330 (MK496766) and isolate Cap3_SP_441 (MK496790), sharing 98% amino acid identity. The two most diverse sequences at the MCP level are from the isolates Cap3_SP_410 (MK496783) and Cap3_SP_433 (MK496788), sharing 16% amino acid identity. A summary of the top 5 BLASTp hits of the MCPs of the capybara associated microviruses from this study is provided in Supplementary Table S3.

4. Conclusions

The application of high throughput sequencing technology in viral metagenomics has allowed the identification of known and novel viruses in a variety of different environments. The number of CRESS DNA viruses identified to date is attributed to viral metagenomic approaches and demonstrates their ubiquity in nature. In this study, we report CRESS DNA viruses ($n = 201$) identified in two faecal samples of capybaras from Brazil. From the full virus genomes identified, 14 belong to the family *Genomoviridae*, one new species is in the family *Smacoviridae*, one new species is from the family *Circoviridae*, and 37 are unclassified diverse CRESS viruses. In addition, 148 novel microviruses were identified. Since the genomes were recovered from faeces, in the case of the prokaryotic-infecting microviruses and smacoviruses, they are likely related to the microbiota of the capybara. The correlation with a host for most CRESS viruses identified through metagenomics approach is a limitation. Nonetheless, this study expands the knowledge of viruses associated with capybaras and is the first report of single-stranded DNA viruses associated with this animal species. Some of these viruses may be infectious to these animals or associated with their microbiota or diet.

Supplementary Materials: The following are available online at <http://www.mdpi.com/1999-4915/11/8/710/s1>, **Supplementary Data 1.** Pairwise identity analysis inferred using SDT v1.2 [32] of the complete genome, Rep, and CP amino acid sequences of the 14 capybara associated genomoviruses with representative sequences of the family *Genomoviridae*. **Supplementary Data 2.** Pairwise identity analysis inferred using SDT v1.2 [32] of Rep and CP amino acid sequences of the capybara associated cyclovirus 1 with representative species of the family *Circoviridae*. **Supplementary Data 3.** Pairwise identity analysis inferred using SDT v1.2 [32] of Rep and CP amino acid of the capybara associated smacovirus with representative sequences of the family *Smacoviridae*. **Supplementary Data 4.** Pairwise identity analysis inferred using SDT v1.2 [32] of Rep amino acid of the unclassified CRESS DNA viruses recovered in this study. SDT analysis was undertaken for the 11 clusters identified in the network analysis that contained the capybara viruses 1–37 (Figure 4). **Supplementary Data 5.** Pairwise identity analysis inferred using SDT v1.2 [32] of microvirus MPC amino acid sequences from the capybara microviruses recovered in this study ($n=148$). **Supplementary Table S1.** Summary of the viruses isolated in this study divided by the family/group in which they are classified with their accession numbers and genome orientation. **Supplementary Table S2.** Rep RCR endonuclease (Motifs I–III) and SF3 helicase motifs (Walker A, Walker B and Motif C) of cyclovirus, genomoviruses, smacovirus and unclassified CRESS DNA viruses identified in this study. **Supplementary Table S3.** Top 5 BLASTx results for MCPs of capybara associated microviruses compared against all the MCPs of all microviruses available in GenBank. **Supplementary Figure S1.** Genome organization of the Capybara associated genomoviruses presented in this study ($n = 14$). **Supplementary Figure S2.** Genome organization of the unclassified CRESS DNA viruses named Capybara virus (1–37) presented in this study. **Supplementary Figure S3.** Genome organization of the Capybara associated microviruses presented in this study ($n = 148$).

Author Contributions: Conceptualization, R.S.F., C.L., A.V. and S.G.R.; methodology, R.S.F., C.L., N.S.L., K.S., A.V. and S.G.R.; formal analysis, R.S.F. and A.V.; investigation, R.S.F., C.L., N.S.L., K.S., A.V. and S.G.R.; resources, C.L., A.V. and S.G.R.; data curation, R.S.F., S.G.R. and A.V.; writing—original draft preparation, R.S.F., and A.V.; writing—review and editing, R.S.F., C.L., N.S.L., K.S., A.V. and S.G.R.; visualization, R.S.F. and A.V.; supervision, A.V. and S.G.R.; project administration, A.V. and S.G.R.; funding acquisition, A.V. and S.G.R.

Funding: This research was partially funded by Embrapa and CNPq. RSF and NSL were recipients of CNPq fellowships.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Pereira, H.D.F.A.; Eston, M. Biologia e manejo de capivaras (*Hydrochoerus hydrochaeris*) no parque estadual Alberto Lófgren, São Paulo, Brasil. *Rev. Inst. Florest.* **2007**, *19*, 55–64.

2. Ferraz, K.M.P.M.B.; Peterson, A.T.; Scachetti-Pereira, R.; Vettorazzi, C.A.; Verdade, L.M. Distribution of Capybaras in an Agroecosystem, Southeastern Brazil, Based on Ecological Niche Modeling. *J. Mammal.* **2009**, *90*, 189–194. [[CrossRef](#)]
3. Mones, A.; Ojasti, J. *Hydrochoerus hydrochaeris*. *Mamm. Species* **1986**, *264*, 1–7. [[CrossRef](#)]
4. De Barros Ferraz, K.M.P.; de Barros Ferraz, S.F.; Moreira, J.R.; Couto, H.T.Z.; Verdade, L.M. Capybara (*Hydrochoerus hydrochaeris*) distribution in agroecosystems: A cross-scale habitat analysis. *J. Biogeogr.* **2007**, *34*, 223–230. [[CrossRef](#)]
5. Verdade, L.M.; Ferraz, K.M. Capybaras in an anthropogenic habitat in Southeastern Brazil. *Braz. J. Biol.* **2006**, *66*, 371–378. [[CrossRef](#)]
6. Krawczak, F.S.; Nieri-Bastos, F.A.; Nunes, F.P.; Soares, J.F.; Moraes-Filho, J.; Labruna, M.B. Rickettsial infection in Amblyomma cajennense ticks and capybaras (*Hydrochoerus hydrochaeris*) in a Brazilian spotted fever-endemic area. *Parasites Vectors* **2014**, *7*, 7. [[CrossRef](#)]
7. Dos SP Andrade, A.C.; Arantes, T.S.; Rodrigues, R.A.; Machado, T.B.; Dornas, F.P.; Landell, M.F.; Furst, C.; Borges, L.G.; Dutra, L.A.; Almeida, G. Ubiquitous giants: A plethora of giant viruses found in Brazil and Antarctica. *Virol. J.* **2018**, *15*, 22. [[CrossRef](#)]
8. Dutra, L.A.; de Freitas Almeida, G.M.; Oliveira, G.P.; Abrahao, J.S.; Kroon, E.G.; Trindade, G.S. Molecular evidence of Orthopoxvirus DNA in capybara (*Hydrochoerus hydrochaeris*) stool samples. *Arch. Virol.* **2017**, *162*, 439–448. [[CrossRef](#)]
9. Barbosa, A.V.; Medaglia, M.L.; Soares, H.S.; Quixabeira-Santos, J.C.; Gennari, S.M.; Damaso, C.R. Presence of neutralizing antibodies to Orthopoxvirus in capybaras (*Hydrochoerus hydrochaeris*) in Brazil. *J. Infect. Dev. Ctries* **2014**, *8*, 1646–1649. [[CrossRef](#)]
10. Moreira, C.A. Notes on the vaccinia virus evolution in wild animals from the Brazilian fauna. *Memórias Inst. Oswaldo Cruz* **1956**, *54*, 93–94. [[CrossRef](#)]
11. Lima, M.T.; Oliveira, G.P.; Afonso, J.A.B.; Souto, R.J.C.; de Mendonca, C.L.; Dantas, A.F.M.; Abrahao, J.S.; Kroon, E.G. An Update on the Known Host Range of the Brazilian Vaccinia Virus: An Outbreak in Buffalo Calves. *Front. Microbiol.* **2018**, *9*, 3327. [[CrossRef](#)]
12. Chiacchio, R.G.; Prioste, F.E.; Vanstreels, R.E.; Knobl, T.; Kolber, M.; Miyashiro, S.I.; Matushima, E.R. Health evaluation and survey of zoonotic pathogens in free-ranging capybaras (*Hydrochoerus hydrochaeris*). *J. Wildl. Dis.* **2014**, *50*, 496–504. [[CrossRef](#)]
13. Catroxo, M.H.B.; Araújo, L.B.D.M.; Lavorenti, A.; Petrella, S.M.C.N.; Melo, N.A.; Martins, A.M.C.R.P. Detection of Coronavirus in Capybaras (*Hydrochoeris hydrochaeris*) by Transmission Electron Microscopy in São Paulo, Brazil. *Int. J. Morphol.* **2010**, *28*, 549–555. [[CrossRef](#)]
14. Kazlauskas, D.; Dayaram, A.; Kraberger, S.; Goldstien, S.; Varsani, A.; Krupovic, M. Evolutionary history of ssDNA bacilladnaviruses features horizontal acquisition of the capsid gene from ssRNA nodaviruses. *Virology* **2017**, *504*, 114–121. [[CrossRef](#)]
15. Breitbart, M.; Delwart, E.; Rosario, K.; Segales, J.; Varsani, A. ICTV Report Consortium. ICTV Virus Taxonomy Profile: Circoviridae. *J. Gen. Virol.* **2017**, *98*, 1997–1998.
16. Zerbini, F.M.; Briddon, R.W.; Idris, A.; Martin, D.P.; Moriones, E.; Navas-Castillo, J.; Rivera-Bustamante, R.; Roumagnac, P.; Varsani, A. ICTV Report Consortium. ICTV Virus Taxonomy Profile: Geminiviridae. *J. Gen. Virol.* **2017**, *98*, 131–133.
17. Krupovic, M.; Ghabrial, S.A.; Jiang, D.; Varsani, A. Genomoviridae: A new family of widespread single-stranded DNA viruses. *Arch. Virol.* **2016**, *161*, 2633–2643. [[CrossRef](#)]
18. Vetten, H.J.; Dale, J.L.; Grigoras, I.; Gronenborn, B.; Harding, R.; Randles, J.W.; Sano, Y.; Thomas, J.E.; Timchenko, T.; Yeh, H.H. Family—Nanoviridae. In *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*; King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J., Eds.; Elsevier: San Diego, CA, USA, 2012; pp. 395–404.
19. Cherwa, J.E.J.; Fane, B.A. Microviridae. In *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*; King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J., Eds.; Elsevier: San Diego, CA, USA, 2012; pp. 385–393.
20. Day, L.A. Inoviridae. In *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*; King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J., Eds.; Elsevier: San Diego, CA, USA, 2012; pp. 375–383.

21. Bamford, D.H.; Pietila, M.K.; Roine, E.; Atanasova, N.S.; Dienstbier, A.; Oksanen, H.M.; Ictv Report, C. ICTV Virus Taxonomy Profile: Pleolipoviridae. *J. Gen. Virol.* **2017**, *98*, 2916–2917. [[CrossRef](#)]
22. Varsani, A.; Krupovic, M. Smacoviridae: A new family of animal-associated single-stranded DNA viruses. *Arch. Virol.* **2018**, *163*, 2005–2015. [[CrossRef](#)]
23. Diez-Villasenor, C.; Rodriguez-Valera, F. CRISPR analysis suggests that small circular single-stranded DNA smacoviruses infect Archaea instead of humans. *Nat. Commun.* **2019**, *10*, 294. [[CrossRef](#)]
24. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D.; et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [[CrossRef](#)]
25. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)]
26. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [[CrossRef](#)]
27. Li, H.; Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **2010**, *26*, 589–595. [[CrossRef](#)]
28. Huang, Y.; Niu, B.; Gao, Y.; Fu, L.; Li, W. CD-HIT Suite: A web server for clustering and comparing biological sequences. *Bioinformatics* **2010**, *26*, 680–682. [[CrossRef](#)]
29. Gerlt, J.A.; Bouvier, J.T.; Davidson, D.B.; Imker, H.J.; Sadkhin, B.; Slater, D.R.; Whalen, K.L. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochim. Biophys. Acta* **2015**, *1854*, 1019–1037. [[CrossRef](#)]
30. Zallot, R.; Oberg, N.O.; Gerlt, J.A. ‘Democratized’ genomic enzymology web tools for functional assignment. *Curr. Opin. Chem. Biol.* **2018**, *47*, 77–85. [[CrossRef](#)]
31. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [[CrossRef](#)]
32. Muhire, B.M.; Varsani, A.; Martin, D.P. SDT: A virus classification tool based on pairwise sequence alignment and identity calculation. *PLoS ONE* **2014**, *9*, e108277. [[CrossRef](#)]
33. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [[CrossRef](#)]
34. Guindon, S.; Dufayard, J.F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **2010**, *59*, 307–321. [[CrossRef](#)]
35. Darriba, D.; Taboada, G.L.; Doallo, R.; Posada, D. ProtTest 3: Fast selection of best-fit models of protein evolution. *Bioinformatics* **2011**, *27*, 1164–1165. [[CrossRef](#)]
36. Stover, B.C.; Muller, K.F. TreeGraph 2: Combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinform.* **2010**, *11*, 7. [[CrossRef](#)]
37. 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. [[CrossRef](#)]
38. Yu, X.; Li, B.; Fu, Y.; Jiang, D.; Ghabrial, S.A.; Li, G.; Peng, Y.; Xie, J.; Cheng, J.; Huang, J.; et al. A geminivirus-related DNA mycovirus that confers hypovirulence to a plant pathogenic fungus. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 8387–8392. [[CrossRef](#)]
39. Kraberger, S.; Arguello-Astorga, G.R.; Greenfield, L.G.; Galilee, C.; Law, D.; Martin, D.P.; Varsani, A. Characterisation of a diverse range of circular replication-associated protein encoding DNA viruses recovered from a sewage treatment oxidation pond. *Infect. Genet. Evol.* **2015**, *31*, 73–86. [[CrossRef](#)]
40. Kraberger, S.; Polston, J.E.; Capobianco, H.M.; Alcalá-Briseno, R.I.; Fontenele, R.S.; Varsani, A. Genomovirus Genomes Recovered from *Echinothrips americanus* Sampled in Florida, USA. *Genome Announc.* **2017**, *5*, e00445-17. [[CrossRef](#)]
41. Rosario, K.; Dayaram, A.; Marinov, M.; Ware, J.; Kraberger, S.; Stainton, D.; Breitbart, M.; Varsani, A. Diverse circular ssDNA viruses discovered in dragonflies (Odonata: Epiprocta). *J. Gen. Virol.* **2012**, *93*, 2668–2681. [[CrossRef](#)]

42. Nakasu, E.Y.; Melo, F.L.; Michereff-Filho, M.; Nagata, T.; Ribeiro, B.M.; Ribeiro, S.G.; Lacorte, C.; Inoue-Nagata, A.K. Discovery of two small circular ssDNA viruses associated with the whitefly *Bemisia tabaci*. *Arch. Virol.* **2017**, *162*, 2835–2838. [[CrossRef](#)]
43. Male, M.F.; 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. [[CrossRef](#)]
44. Schmidlin, K.; Sepp, T.; Khalifeh, A.; Smith, K.; Fontenele, R.S.; McGraw, K.J.; Varsani, A. Diverse genomoviruses representing eight new and one known species identified in feces and nests of house finches (*Haemorrhous mexicanus*). *Arch. Virol.* **2019**. [[CrossRef](#)]
45. Rosario, K.; Breitbart, M.; Harrach, B.; Segales, J.; Delwart, E.; Biagini, P.; Varsani, A. Revisiting the taxonomy of the family Circoviridae: Establishment of the genus Cyclovirus and removal of the genus Gyrovirus. *Arch. Virol.* **2017**, *162*, 1447–1463. [[CrossRef](#)]
46. Ge, X.; Li, J.; Peng, C.; Wu, L.; Yang, X.; Wu, Y.; Zhang, Y.; Shi, Z. Genetic diversity of novel circular ssDNA viruses in bats in China. *J. Gen. Virol.* **2011**, *92*, 2646–2653. [[CrossRef](#)]
47. Li, L.L.; Kapoor, A.; Slikas, B.; Bamidele, O.S.; Wang, C.L.; Shaukat, S.; Masroor, M.A.; Wilson, M.L.; Ndjanga, J.B.N.; Peeters, M.; et al. Multiple Diverse Circoviruses Infect Farm Animals and Are Commonly Found in Human and Chimpanzee Feces. *J. Virol.* **2010**, *84*, 1674–1682. [[CrossRef](#)]
48. Zhang, W.; Li, L.; Deng, X.; Kapusinszky, B.; Pesavento, P.A.; Delwart, E. Faecal virome of cats in an animal shelter. *J. Gen. Virol.* **2014**, *95*, 2553–2564. [[CrossRef](#)]
49. De Sales Lima, F.E.; Cibulski, S.P.; dos Santos, H.F.; Teixeira, T.F.; Varela, A.P.M.; Roehle, P.M.; Delwart, E.; Franco, A.C. Genomic characterization of novel circular ssDNA viruses from insectivorous bats in Southern Brazil. *PLoS ONE* **2015**, *10*, e0118070.
50. Steel, O.; Kraberger, S.; Sikorski, A.; Young, L.M.; Catchpole, R.J.; Stevens, A.J.; Ladley, J.J.; Coray, D.S.; Stainton, D.; Dayaram, A.; et al. Circular replication-associated protein encoding DNA viruses identified in the faecal matter of various animals in New Zealand. *Infect. Genet. Evol.* **2016**, *43*, 151–164. [[CrossRef](#)]
51. Kapusinszky, B.; Ardeshir, A.; Mulvaney, U.; Deng, X.; Delwart, E. Case-Control Comparison of Enteric Viromes in Captive Rhesus Macaques with Acute or Idiopathic Chronic Diarrhea. *J. Virol.* **2017**, *91*, e00952-17. [[CrossRef](#)]
52. Reuter, G.; Boros, A.; Delwart, E.; Pankovics, P. Novel circular single-stranded DNA virus from turkey faeces. *Arch. Virol.* **2014**, *159*, 2161–2164. [[CrossRef](#)]
53. Lima, D.A.; Cibulski, S.P.; Tochetto, C.; Varela, A.P.M.; Finkler, F.; Teixeira, T.F.; Loiko, M.R.; Cerva, C.; Junqueira, D.M.; Mayer, F.Q.; et al. The intestinal virome of malabsorption syndrome-affected and unaffected broilers through shotgun metagenomics. *Virus Res.* **2019**, *261*, 9–20. [[CrossRef](#)]
54. Tisza, M.J.; Pastrana, D.V.; Welch, N.L.; Stewart, B.; Peretti, A.; Starrett, G.J.; Pang, Y.-Y.S.; Varsani, A.; Krishnamurthy, S.R.; Pesavento, P.A. Discovery of several thousand highly diverse circular DNA viruses. *BioRxiv* **2019**. [[CrossRef](#)]
55. Pearson, V.M.; Caudle, S.B.; Rokyta, D.R. Viral recombination blurs taxonomic lines: Examination of single-stranded DNA viruses in a wastewater treatment plant. *PeerJ* **2016**, *4*, e2585. [[CrossRef](#)]
56. Rosario, K.; Schenck, R.O.; Harbeitner, R.C.; Lawler, S.N.; Breitbart, M. Novel circular single-stranded DNA viruses identified in marine invertebrates reveal high sequence diversity and consistent predicted intrinsic disorder patterns within putative structural proteins. *Front. Microbiol.* **2015**, *6*, 696. [[CrossRef](#)]
57. Dayaram, A.; Galatowitsch, M.L.; Arguello-Astorga, G.R.; van Bysterveldt, K.; Kraberger, S.; Stainton, D.; Harding, J.S.; Roumagnac, P.; Martin, D.P.; Lefevre, P.; et al. Diverse circular replication-associated protein encoding viruses circulating in invertebrates within a lake ecosystem. *Infect. Genet. Evol.* **2016**, *39*, 304–316. [[CrossRef](#)]
58. Kraberger, S.; Schmidlin, K.; Fontenele, R.S.; Walters, M.; Varsani, A. Unravelling the Single-Stranded DNA Virome of the New Zealand Blackfly. *Viruses* **2019**, *11*, 532. [[CrossRef](#)]
59. Kraberger, S.; Cook, C.N.; Schmidlin, K.; Fontenele, R.S.; Bautista, J.; Smith, B.; Varsani, A. Diverse single-stranded DNA viruses associated with honey bees (*Apis mellifera*). *Infect. Genet. Evol.* **2019**, *71*, 179–188. [[CrossRef](#)]
60. Dayaram, A.; Potter, K.A.; Pailes, R.; Marinov, M.; Rosenstein, D.D.; Varsani, A. Identification of diverse circular single-stranded DNA viruses in adult dragonflies and damselflies (Insecta: Odonata) of Arizona and Oklahoma, USA. *Infect. Genet. Evol.* **2015**, *30*, 278–287. [[CrossRef](#)]

61. Roux, S.; Krupovic, M.; Poulet, A.; Debroas, D.; Enault, F. Evolution and diversity of the Microviridae viral family through a collection of 81 new complete genomes assembled from virome reads. *PLoS ONE* **2012**, *7*, e40418. [[CrossRef](#)]
62. Walters, M.; Bawuro, M.; Christopher, A.; Knight, A.; Kraberger, S.; Stainton, D.; Chapman, H.; Varsani, A. Novel Single-Stranded DNA Virus Genomes Recovered from Chimpanzee Feces Sampled from the Mambilla Plateau in Nigeria. *Genome Announc.* **2017**, *5*, e01715-16. [[CrossRef](#)]
63. Creasy, A.; Rosario, K.; Leigh, B.A.; Dishaw, L.J.; Breitbart, M. Unprecedented Diversity of ssDNA Phages from the Family Microviridae Detected within the Gut of a Protochordate Model Organism (*Ciona robusta*). *Viruses* **2018**, *10*, 404. [[CrossRef](#)]
64. Hopkins, M.; Kailasan, S.; Cohen, A.; Roux, S.; Tucker, K.P.; Shevenell, A.; Agbandje-McKenna, M.; Breitbart, M. Diversity of environmental single-stranded DNA phages revealed by PCR amplification of the partial major capsid protein. *ISME J.* **2014**, *8*, 2093–2103. [[CrossRef](#)]
65. Angly, F.E.; Felts, B.; Breitbart, M.; Salamon, P.; Edwards, R.A.; Carlson, C.; Chan, A.M.; Haynes, M.; Kelley, S.; Liu, H.; et al. The marine viromes of four oceanic regions. *PLoS Biol.* **2006**, *4*, e368. [[CrossRef](#)]
66. Labonte, J.M.; Suttle, C.A. Metagenomic and whole-genome analysis reveals new lineages of gokushoviruses and biogeographic separation in the sea. *Front. Microbiol.* **2013**, *4*, 404. [[CrossRef](#)]
67. Lopez-Bueno, A.; Tamames, J.; Velazquez, D.; Moya, A.; Quesada, A.; Alcami, A. High diversity of the viral community from an Antarctic lake. *Science* **2009**, *326*, 858–861. [[CrossRef](#)]
68. Krupovic, M.; Dutilh, B.E.; Adriaenssens, E.M.; Wittmann, J.; Vognesen, F.K.; Sullivan, M.B.; Rumnies, J.; Prangishvili, D.; Lavigne, R.; Kropinski, A.M.; et al. Taxonomy of prokaryotic viruses: Update from the ICTV bacterial and archaeal viruses subcommittee. *Arch. Virol.* **2016**, *161*, 1095–1099. [[CrossRef](#)]
69. Brentlinger, K.L.; Hafenstein, S.; Novak, C.R.; Fane, B.A.; Borgon, R.; McKenna, R.; Agbandje-McKenna, M. Microviridae, a family divided: Isolation, characterization, and genome sequence of ϕ MH2K, a bacteriophage of the obligate intracellular parasitic bacterium *Bdellovibrio bacteriovorus*. *J. Bacteriol.* **2002**, *184*, 1089–1094. [[CrossRef](#)]
70. Garner, S.A.; Everson, J.S.; Lambden, P.R.; Fane, B.A.; Clarke, I.N. Isolation, molecular characterisation and genome sequence of a bacteriophage (Chp3) from *Chlamydomonas reinhardtii*. *Virus Genes* **2004**, *28*, 207–214. [[CrossRef](#)]
71. Chipman, P.R.; Agbandje-McKenna, M.; Renaudin, J.; Baker, T.S.; McKenna, R. Structural analysis of the spiroplasma virus, SpV4: Implications for evolutionary variation to obtain host diversity among the Microviridae. *Structure* **1998**, *6*, 135–145. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).