Virology 554 (2021) 89–96

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/virology

# Diverse cressdnaviruses and an anellovirus identified in the fecal samples of yellow-bellied marmots

Anthony Khalifeh [a], Daniel T. Blumstein [b,**], Rafaela S. Fontenele [a], Kara Schmidlin [a], Cécile Richet [a], Simona Kraberger [a], Arvind Varsani [a,c,*]

[a] The Biodesign Center for Fundamental and Applied Microbiomics, School of Life Sciences, Center for Evolution and Medicine, Arizona State University, Tempe, AZ, 85287, USA
[b] Department of Ecology & Evolutionary Biology, Institute of the Environment & Sustainability, University of California Los Angeles, Los Angeles, CA, 90095, USA
[c] Structural Biology Research Unit, Department of Clinical Laboratory Sciences, University of Cape Town, 7925, Cape Town, South Africa

ARTICLE INFO

*Keywords:*
Marmota flaviventer
Anelloviridae
Genomoviridae
Cressdnaviricota

ABSTRACT

Over that last decade, coupling multiple strand displacement approaches with high throughput sequencing have resulted in the identification of genomes of diverse groups of small circular DNA viruses. Using a similar approach but with recovery of complete genomes by PCR, we identified a diverse group of single-stranded viruses in yellow-bellied marmot (*Marmota flaviventer*) fecal samples. From 13 fecal samples we identified viruses in the family *Genomoviridae* (n = 7) and *Anelloviridae* (n = 1), and several others that ware part of the larger *Cressdnaviricota* phylum but not within established families (n = 19). There were also circular DNA molecules identified (n = 4) that appear to encode one viral-like gene and have genomes of <1545 nts. This study gives a snapshot of viruses associated with marmots based on fecal sampling.

## 1. Introduction

An increasing number of novel circular single-stranded DNA viruses have been discovered in the last decade. This can be largely attributed to the innovations and advancements in metagenomic sequencing, in particular coupling of multiple strand displace amplification and high throughput. A large portion of these viruses are part of recently established phylum *Cressdnaviricota* (Krupovic et al., 2020). Cressdnaviruses range from ~1 kb to 6kb and encode between 2 and 10 opening reading frames (ORFs) (Rosario et al., 2012; Zhao et al., 2019). Cressdnaviruses encode a distinct replication associated protein (Rep) that has conserved motifs, an origin of replication, and a capsid protein. There are currently six established eukaryotic cressdnavirus families: *Bacilladnaviridae*, *Circoviridae*, *Geminiviridae*, *Genomoviridae*, *Nanoviridae*, *Redondoviridae* and *Smacoviridae*. There is a large number of viruses within the *Cressdnaviricota* phylum that do not fall into any currently established families. Additionally, other small circular ssDNA viruses which do not encode a replication-associated protein are classified in the families *Anelloviridae*, *Inoviridae*, *Microviridae* and *Spiraviridae*.

Previous studies have described a handful of viruses in marmots in genus Marmota (Armitage, 2014), including California encephalitis virus (ssRNA) that has been found in yellow-bellied marmots (McLean et al., 1968), rabies virus (ssRNA) in groundhogs (*Marmota monax*) (Blanton et al., 2011) and, more recently a bocaparvovirus (ssDNA) has been identified in Himalayan marmots (*Marmota himalayana*) (Ao et al., 2017). Yellow-bellied marmots (*Marmota flaviventer*) are one of the 15 species in the. Yellow-bellied marmots are semi-fossorial, ground-dwelling, sciurid rodents native to the western United States of America. They are facultatively social and live in colonies that range from a few to >50 individuals. Yellow-bellied marmots are 3–5 kg facultatively social, diurnal, semi-fossorial, sciurid rodents found in the mountains and intermountain regions of western North America (Armitage, 2014; Frase and Hoffmann, 1980). Their burrows provide protection from predators and a place in which they hibernate for 7–8 months per year. Colony sites may contain one to several adult females, one to several adult males, plus yearlings. Many yearlings disperse before the young of the year emerge above ground. These herbivores typically double their body mass during their summer active season (Heissenberger et al., 2020),

* Corresponding author. , The Biodesign Center for Fundamental and Applied Microbiomics, School of Life Sciences, Center for Evolution and Medicine, Arizona State University, Tempe, AZ, 85287, USA.
** Corresponding author.
*E-mail addresses:* marmots@ucla.edu (D.T. Blumstein), Arvind.Varsani@asu.edu (A. Varsani).

https://doi.org/10.1016/j.virol.2020.12.017
Received 19 September 2020; Received in revised form 24 December 2020; Accepted 25 December 2020
Available online 30 December 2020

and a growing literature (e.g. (Maldonado-Chaparro et al., 2018; Ozgul et al., 2010),) have studied factors associated with this. Relatively little is known about viruses associated with any species of marmot.

In this study we primarily identified viruses with similarities to circular ssDNA viruses. Viruses with similarities to members of the *Genomoviridae* and *Anelloviridae* families were identified along with those similar to other unclassified CRESS DNA viruses which fall within the phylum *Cressdnaviricota* (Krupovic et al., 2020).

## 2. Material and methods

### 2.1. Sample collection, viral nucleic acid isolation and high-throughput sequencing

Fresh fecal samples were collected from 35 yellow-bellied marmots living in five colonies located sampled in the upper East River Valley, in and around the Rocky Mountain Biological Laboratory near Crested Butte, Colorado, the site of a long-term (59-year) study (Armitage, 2014; Blumstein, 2013). Samples were collected in May and June 2018 within 3 km of each other. 5 g of each fecal sample was homogenized by vortexing in 10 ml SM buffer (0.1 M NaCl, 50 mM Tris/HCl-pH 7.4, and 10 mM $MgSO_4$) and processed as outline in Steel et al. (2016). The viral DNA was extracted using High Pure viral nucleic acid kit (Roche Diagnostics, USA) according to manufacturer's instructions and circular DNA was amplified using rolling circle amplification (RCA) with TempliPhi 2000 kit (GE Healthcare, USA). The RCA products were pooled and used to prepare Illumina sequencing libraries which were sequenced on the Illumina 4000 platform at Macrogen Inc. (Korea).

The raw reads were *de novo* assembled using SPAdes v 3.12.0 (Bankevich et al., 2012) and the resulting contigs (>1000 nts) were analyzed using BLASTx (Altschul et al., 1990) against a viral protein database. The contigs were sorted based upon similarities to viral families. For this study we decided to focus on recovering compete genomes and thus we focused on small circular DNA viruses. Abutting primers (Supplementary Table 1) were designed based on these circular ssDNA virus-like contigs and these were used to amplify the viral sequences using PCR with Kapa HiFi Hotstart DNA polymerase (Kapa Biosystems, USA) following the manufacturer recommend thermal cycling conditions with an annealing temperature of 55 °C and 60 °C. Amplified genomes were resolved on a 0.7% agarose gel, excised from the gel and purified. The purified amplicons were ligated with the pJET 1.2 vector (Thermo Fisher Scientific, USA) and transformed into *Escherichia coli* DH5-alpha competent cells. The purified recombinant plasmids were Sanger sequenced at Macrogen Inc. (Korea) by primer walking. The Sanger sequence contigs were assembled and annotated using Geneious v11.0.3. The sequences (n = 31) of the viral genomes/molecules recovered in this study were deposited in the GenBank database (Accession numbers: MT044319–MT044326, MT181524–MT181546).

## 3. Sequence similarity network analyses

### 3.1. Anellovirus sequence analyses

Since the largest open reading frame (ORF) is relatively conserved in all anelloviruses, we assembled a dataset of all anellovirus ORF1 available in GenBank (downloaded on the 1st June 2020). The ORF1 protein sequences were aligned using MUSCLE (Edgar, 2004). The resulting ORF1 alignment was used to infer a maximum likelihood phylogeny using PHYML (Guindon et al., 2010) with the substitution model rtREV + G + F inferred from using ProtTest (Darriba et al., 2011) as the best fit model. Branch support <0.8 aLRT support were collapsed using TreeGraph2 (Stover and Muller, 2010).

### 3.2. Genomovirus sequences analyses

The Rep protein sequences from the seven marmot-derived genomoviruses together with those encoded by genomovirus genomes sequences available in GenBank (n = 468) were aligned with MUSCLE (Edgar, 2004) and the resulting alignment was used to construct a Maximum Likelihood phylogenetic tree with FastTree (Price et al., 2010) using WAG + G amino acid substitution model. The tree was rooted with the sequences of geminiviruses. Branch support <0.8 aLRT support were collapsed using TreeGraph2 (Stover and Muller, 2010).

### 3.3. Unclassified cressdnaviruses and circular DNA molecules

Datasets of the Rep proteins of unclassified CRESS DNA viruses identified in this study were combined with those available in GenBank into a dataset. Rep protein sequences of these viruses together with the ones from this study were used to generate a sequence similarity networks using EST-EFI (Gerlt et al., 2015). We used a network threshold of 60 as from our previous experience (Fontenele et al., 2019; Orton et al., 2020) this allows for viral family level clustering of the classified viruses. The networks were visualized using Cytoscape V3.7.1 (Shannon et al., 2003) with the organic layout. The network clusters containing Reps of unclassified viruses from this study (n = 5) were extracted and used to infer Maximum Likelihood phylogenies. The extracted sequences were aligned with MUSCLE (Edgar, 2004) and this alignment was used to infer a Maximum Likelihood phylogenetic tree using PHYML (Guindon et al., 2010) with WAG + G + I amino acid substitution model inferred from using ProtTest (Darriba et al., 2011) as the best fit model. The trees were midpoint rooted. Branch support <0.8 aLRT support were collapsed using TreeGraph2 (Stover and Muller, 2010).

### 3.4. Pairwise identities

All genome-wide and protein specific pairwise identities were determined using SDT 1.2 (Muhire et al., 2014).

## 4. Results and discussion

We identified 15 contigs with similarities to anelloviruses, genomoviruses and unclassified cressdnaviruses. Based on these contigs, we designed 15 pairs of abutting primers (Supplementary Table 1) to screen and recover the complete genomes of these circular viruses from all the fecal samples.

We recovered a virus that belongs to the family *Anelloviridae*, seven to the family *Genomoviridae*, 19 are part of the larger unclassified circular replication associated encoding single-stranded (CRESS) DNA virus groups within the phylum *Cressdnaviricota* (Krupovic et al., 2020) and four circular DNA molecules (encoding either a Rep, CP or viral-like ORF). A summary of the recovered viruses and their genome organization is provided in Fig. 1. For the purpose of this study, we used genotypes threshold of 98% based on genome-wide pairwise identity determined using SDT 1.2 (Muhire et al., 2014). Based on this criteria, we have one anellovirus genotype, four of genomoviruses, six of unclassified cressdnaviruses and four of viral-like circular DNA molecules (Fig. 1). For two of the genotypes (MarFaV3 and 4) of the unclassified cressdnaviruses we identified two viral genomes per samples that share >98% pairwise identity. The anellovirus, two genomoviruses (MarFaFmV1 and 3), three cressdnaviruses (MarFaV2, 5 and 6) and the four viral-like circular molecules (MarFACM1-4) were found in only one marmot fecal samples, whereas the other were found in >2 samples with MarFaV4 being identified in 5 samples. The maximum number of viral genotypes identified in a single fecal sample was six (Fig. 1).

### 4.1. Anellovirus

*Anelloviridae* is a family of viruses that are known to infect mammal and avian species (Cibulski et al., 2014; Crane et al., 2018; de Souza et al., 2018; Hrazdilova et al., 2016; Liu et al., 2011; Nishiyama et al., 2014, 2015). Certain anellovirus species have been shown to have 100%

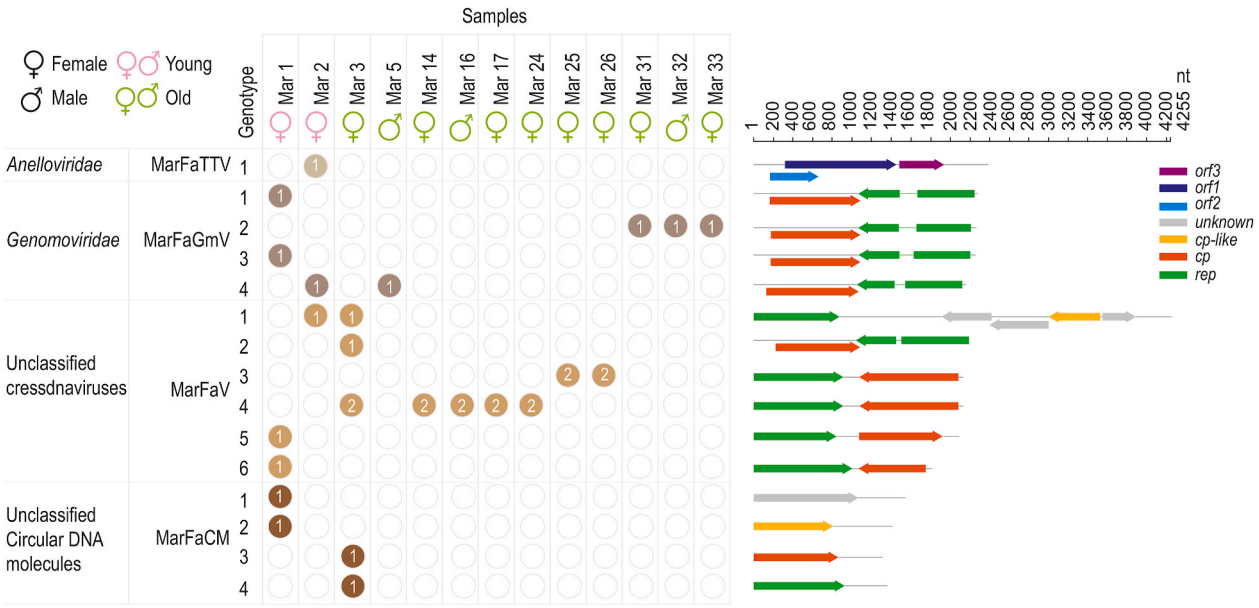


**Fig. 1.** Summary of the viruses identified in 13 individual marmot fecal samples. Colored circles represent virus genotype identified in each virus group and the numbers in associated with each circle represent the number of individual genomes for that particular genotype. A linearized genome organization of the viruses and genotypes is presented on the right.

prevalence in human populations (Freer et al., 2018). They have one ORF, ORF1, and few smaller ORFs. *Anelloviridae* is currently composed of 10 genera, *Alphatorquevirus*, *Betatorquevirus*, *Gammatorquevirus*, *Gyrovirus*, *Deltatorquevirus*, *Epsilontorquevirus*, *Etatorquevirus*, *Iotatorquevirus*, *Thetatorquevirus and Zetatorquevirus,* (Biagini et al., 2011). Viruses in this family viruses are classified based on a 35% nucleotide identity variation of the ORF1. We identified a novel anellovirus referred to as marmot-associated torque teno virus 1 (MarTTV1) (MT044319) from one marmot fecal sample (Table 1) and it is part of a new genus tentatively named Aleptorquevirus (Varsani and Kraberger, 2020). The ORF1 of MarTTV1 is most closely related to two anelloviruses associated with Iberian hare (MN994854 and MN994867)

**Table 1**
Summary of virus and circular molecule sequences recovered in this study and the HUH endonucleases and superfamily 3 helicases identified in the replication-associated proteins cressdnaviruses.

| Family | Genus | Virus | Accession # | Motif I | Motif II | Motif III | Walker A | Walker B | Motif C |
|---|---|---|---|---|---|---|---|---|---|
| *Anelloviridae* | *Aleptorquevirus* | MarFaTTV 1 | MT044319 | - | - | - | - | - | - |
| *Genomoviridae* | *Gemycircularvirus* | MarFaGmV 1 | MT044325 | LLTYAQ | HLHV | EKGYDYAIK | GPSRTGKTTWAR | VFDDV | IWCSN |
| | | MarFaGmV 2 | MT044320 | FLTYAQ | HLHV | EKGYDYAIK | GPSRTGKTSWAR | VFDDM | IWCAN |
| | | MarFaGmV 2 | MT044321 | LLTYAQ | HLHV | EKGYDYAIK | GPSRTGKTSWAR | VFDDM | IWCAN |
| | | MarFaGmV 2 | MT044322 | LLTYAQ | HLHV | EKGYDYAIK | GPSRTGKTSWAR | VFDDM | IWCAN |
| | | MarFaGmV 3 | MT044326 | LLTYAQ | HLHV | EKGYDYAIK | GPSRTGKTSWAR | VFDDM | IWCAN |
| | | MarFaGmV 4 | MT044323 | LVTYSH | HFHV | EAGYDYAVK | GPSRLGKTVWSR | VFDDI | IWVAN |
| | | MarFaGmV 4 | MT044324 | LVTYSH | HFHV | EAGYDYAVK | GPSRLGKTVWSR | VFDDI | IWVAN |
| Unassigned | Unassigned | MarFaV 1 | MT181542 | - | HYHV | VEYVKYCDK | GPAGTGKSRKAF | IIDDW | IVTSN |
| | | MarFaV 1 | MT181544 | - | HYHV | VEYVKYCDK | GPAGTGKSRKAF | IIDDW | IVTSN |
| | | MarFaV 2 | MT181541 | FLTYSQ | HFHV | FNRRHYIRK | GPTLLGKTAFIR | VFDDV | IFICN |
| | | MarFaV 3 | MT181528 | QLTLNE | HIHI | SQNIAYIEK | GASGIGKTEKAK | LYDDF | IITSV |
| | | MarFaV 3 | MT181529 | QLTLNE | HIHI | SQNIAYIEK | GASGIGKTEKAK | LYDDF | IITTM |
| | | MarFaV 3 | MT181530 | QLTLNE | HIHI | SQNIAYIEK | GASGIGKTEKAK | LYDDF | IITSV |
| | | MarFaV 3 | MT181531 | QLTLNE | HIHI | SQNIAYIEK | GASGIGKTEKAK | LYDDF | IITSV |
| | | MarFaV 4 | MT181535 | QLTLNE | HIHI | SQNIAYIEK | GASGIGKTEKAK | LYDDF | IITSV |
| | | MarFaV 4 | MT181536 | QLTLNE | HIHI | SQNIAYIEK | GASGIGKTEKAK | LYDDF | IITSV |
| | | MarFaV 4 | MT181537 | QLTLNE | HIHI | SQNIAYIEK | GASGIGKTEKAK | LYDDF | IITSV |
| | | MarFaV 4 | MT181538 | QLTLNE | HIHI | SQNIAYIEK | GASGIGKTEKAK | LYDDF | IITSV |
| | | MarFaV 4 | MT181539 | QLTLNE | HIHI | SQNIAYIEK | GASGIGKTEKAK | LYDDF | IMTSV |
| | | MarFaV 4 | MT181540 | QLTLNE | HIHI | SQNIAYIEK | GASGIGKTEKAK | LYDDF | IITSV |
| | | MarFaV 4 | MT181543 | QLTLNE | HIHI | SQNIAYIEK | GASGIGKTEKAK | LYDDF | IITSV |
| | | MarFaV 4 | MT181532 | QLTLNE | HIHI | SQNIAYIEK | GASGIGKTEKAK | LYDDF | IITSV |
| | | MarFaV 4 | MT181533 | QLTLNE | HIHI | SQNIAYIEK | GASGIGKTEKAK | LYDDF | IITSV |
| | | MarFaV 4 | MT181534 | QLTLNE | HIHI | SQNIAYIEK | GASGIGKTEKAK | LYDDF | IITSV |
| | | MarFaV 5 | MT181545 | CFTLNN | HLQG | RQNRVYCSK | GRPGVGKSRLAH | IIDDF | IVTSN |
| | | MarFaV6 | MT181546 | VFTFNN | HLQG | KQARDYCLK | GNSGNGKTRLAK | IIDDF | VFTSP |
| Unassigned | Unassigned | MarFaCM 1 | MT181525 | - | - | - | - | - | - |
| | | MarFaCM 2 | MT181526 | - | - | - | - | - | - |
| | | MarFaCM 3 | MT181524 | - | - | - | - | - | - |
| | | MarFaCM 4 | MT181527 | - | - | AQNVKYCLI | GPAGTNKTRSAF | IIDDF | FITSC |

(Águeda-Pinto et al., 2020) sharing 51% amino acid similarity. Based upon the 65% ORF1 nucleotide identity species demarcation coupled with phylogenetic support, MarTTV-1 represents a new species in the new genus (Fig. 2). The phylogenetic analysis of the ORF1 of MarTTV1 show relatedness with rodent anelloviruses, including hare and mosquito-derived anelloviruses. The mosquito anellovirus may seem to be an outlier, but mosquitos feed on blood therefore being able to acquire the rodent strains while carrying the bloodmeal.

### 4.2. Genomoviruses

*Genomoviridae* is a recently established family of circular single-stranded virus (Krupovic et al., 2016). *Genomoviridae* encodes for a Rep (often spliced), and a CP, and is currently comprised of nine genera, *Gemycircularvirus*, *Gemyduguivirus*, *Gemygorvirus*, *Gemykibivirus*, *Gemykolovirus*, *Gemykrogvirus*, *Gemykroznavirus*, *Gemytondivrus* and *Gemyvongvirus*. Nonetheless, only one virus from this family has been shown to infect and repress the pathogenic effects of its host, a fungus (Yu et al., 2010). Viruses in the family *Genomoviridae* have been identified in a variety of environmental samples, as well as plant and animal samples (Fontenele et al., 2019; Kraberger et al., 2018b; Krupovic et al., 2016; Sikorski et al., 2013; Steel et al., 2016).

Based upon 78% full-genome pairwise identity species demarcation threshold (Varsani and Krupovic, 2017) and well-supported phylogenetic clades, these viruses can be classified into four new species referred to as marmot-associated genomovirus 1–4 (MarGV) (MT044320-MT044326), seven variants, (Table 1, Fig. 3). All MarGVs can be assigned to the *Gemycircularvirus* genus and all have the same nonanucleotide sequence "TAATATTAT". The Rep of MarGV1 is most closely related to a Rep encoded by sierra dome spider-associated circular virus (MH545510) (Rosario et al., 2018) sharing 74% amino acid pairwise identity. The Rep of MarGV-2 and MarGV-3 are most closely related to that of a gemycircularvirus isolate 51_Fec80064_sheep (KT862251) (Steel et al., 2016) sharing 91–96% amino acid pairwise identity. The Rep of MarGV4 is most closely related to that of the genomovirus isolate BbaGV-4_US-AB02_249–2014 (MG571100 (Kraberger et al., 2018a); sharing 76% identity.

### 4.3. Cressdnaviruses and circular molecules

An increasing number of novel single-stranded DNA viruses have been discovered in the last decade. Many of these do not fall within established families and are loosely labelled as unclassified CRESS DNA viruses within the recently establish phylum *Cressdnaviricota* (Krupovic et al., 2020). Many of the unclassified CRESS DNA viruses have been found in a wide variety of environments and have been associated with a diverse range of hosts and their genomes range ~1.6 kb - 6 kb and encode between 2 and 10 ORFs (Rosario et al., 2012; Zhao et al., 2019).

Nineteen novel cressdnaviruses were discovered in this study isolated from 9 samples, 16 of which group within 4 network clusters (Table 1, Fig. 4). These viruses range in size from ~1.8–4.2 kb and are referred to as marmot associated feces virus (MarFaV)1–8. Within

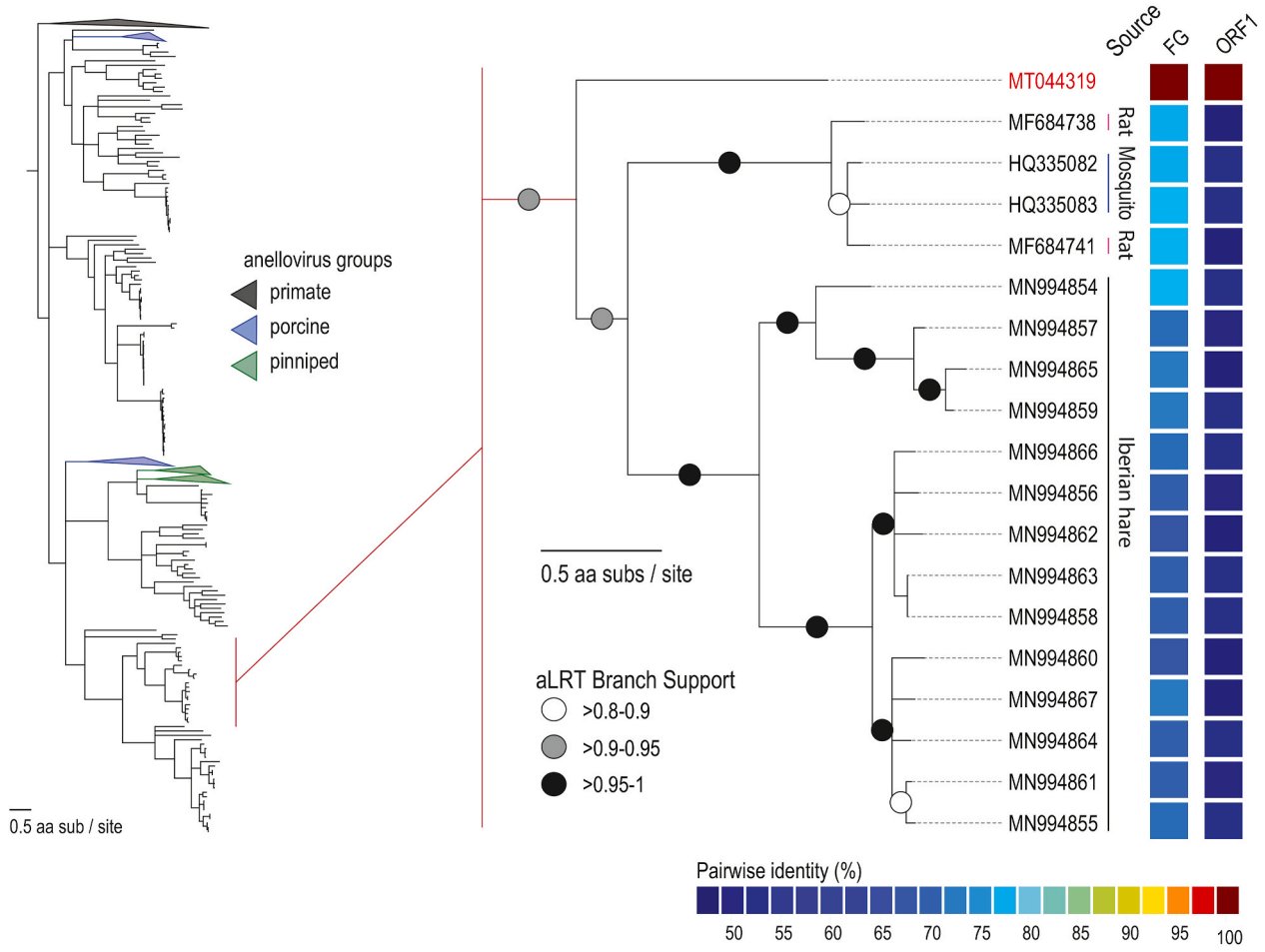


**Fig. 2.** Phylogenetic tree of the aligned ORF1 protein sequence encoded by all complete anellovirus genomes available in GenBank (downloaded on the 1st June 2020) together with the one of the marmot associated anellovirus highlighted in red. A zoomed subtree is provided to the right that shows the members of the proposed new genus Aleptorquevirus and the source of the anelloviruses that are part of this genus (Varsani and Kraberger, 2020). Percentage pairwise comparison of full genome nucleotide and ORF1 amino acid sequence is shown to the right of the aleptorquevirus subtree.

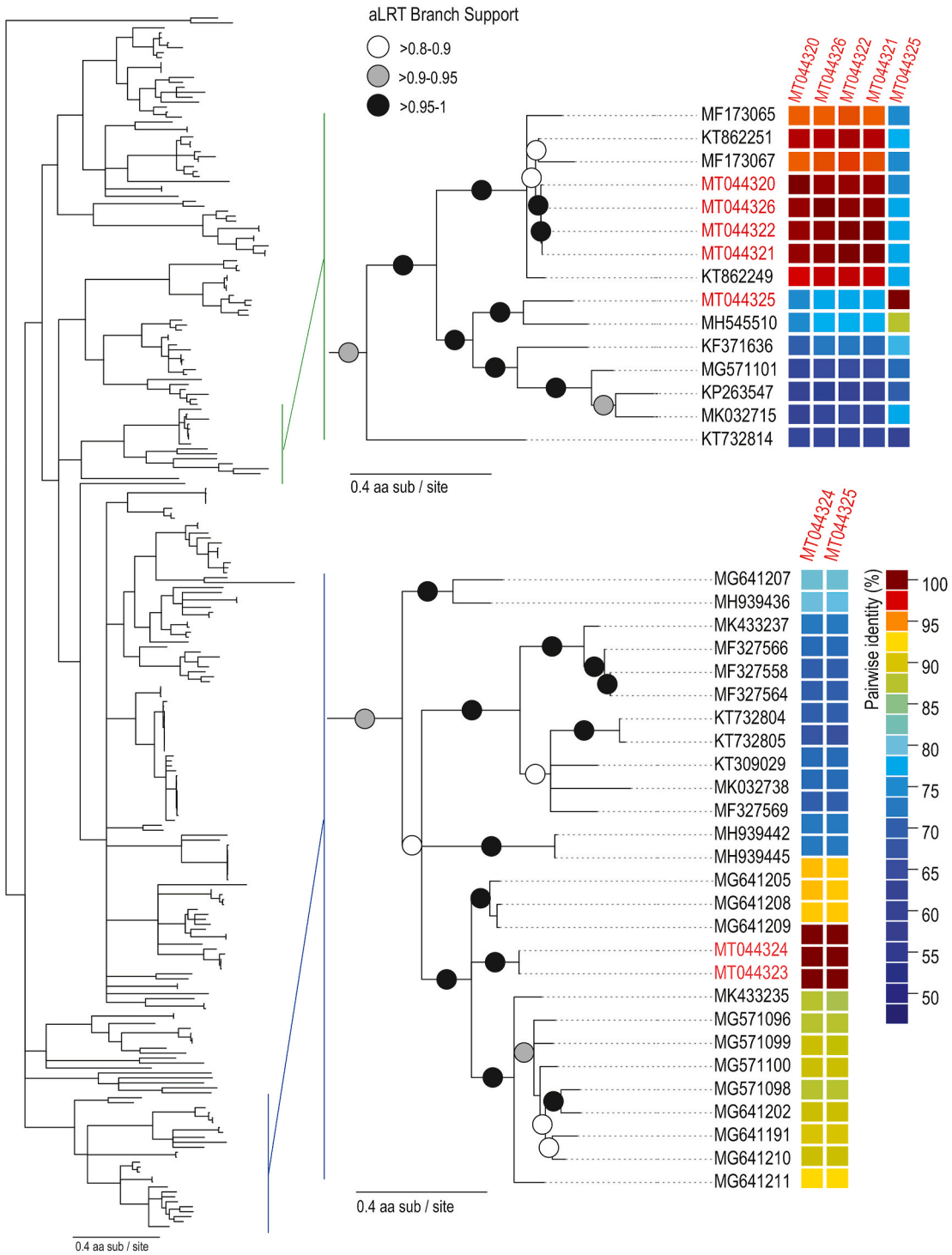


**Fig. 3.** Phylogenetic analysis of genomovirus replication-associated protein (Rep) sequences in the *Gemycircularvirus* genus. Zoomed in section of the two clades with sequences identified in this study are shown. Marmot-associated virus sequences are highlighted in red. Pairwise comparison of the genomovirus full genome, replication-associated protein (Rep) sequence and capsid protein are shown to the right of the phylogenetic tree.

cluster 1 the Rep of MarFaV6 is most closely related to that of a virus from a lake sediment sample (KP153408) sharing 45% amino acid identity. Within cluster 2 the Rep of MarFaV5 is most closely related to that of cressdnavirus isolate BtMl-CV/QH2013 (KJ641730) (Wu et al., 2016), sharing a 78% amino acid identity. Within cluster 3 the Reps of MarFaV3-4 are most closely related that of a cressdnavirus identified in a turkey tissue sample (MK012476) (Tisza et al., 2020), sharing between 43.6 and 44.6% Rep amino acid identity. Several viruses grouped into smaller clusters, such as MarFaV1 and -2. Within cluster 4 the Rep of MarFaV2 is most closely related to that of a sewage associated circular DNA virus (KM821748) (Kraberger et al., 2015), sharing 41% identity. Within Cluster 5 the Reps of both MarFaV-1s are most closely related that of a cressdnavirus identified in a rainbow trout sample (MH617762) (Tisza et al., 2020), both sharing a 42% Rep identity. Rep belonging to MarFaV6 is within a clade containing viruses from an eclectic range of sample sources including cow tissue (MH617295) (Tisza et al., 2020), red snapper tissue (MH616644) (Tisza et al., 2020), New Zealand cockle (KM874315) (Dayaram et al., 2015), and marine sample (JX904076)

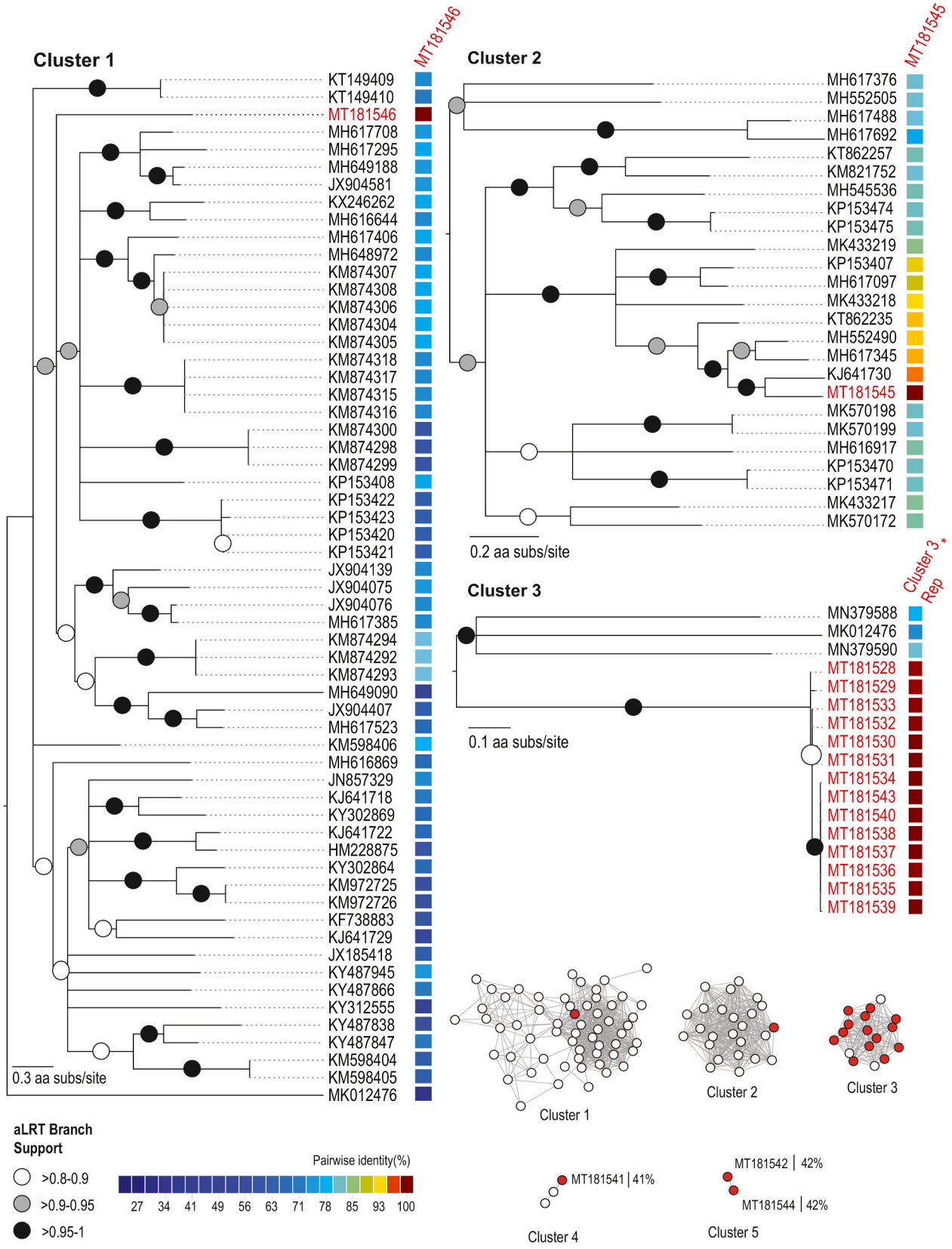


**Fig. 4.** Sequence similarity networks and phylogenetic analysis of replication-associated protein (Rep) within CRESS DNA virus clusters. Marmot-associated virus sequences are highlighted in red. Pairwise comparison of CRESS DNA virus replication-associated protein (Rep) of clusters 1, 2, and 3 are shown.

(Labonte and Suttle, 2013). Rep belonging to MarFV5 clusters with a Rep from bat circovirus (KJ641730) (Wu et al., 2016). Reps of MarFaV3-4 are most closely related to those of viruses from turkey tissue (MK012476) (Tisza et al., 2020) as well as chicken samples (MN379588 & MN379590).

Several circular DNA molecules were also identified in this study. These are referred to as marmot feces-associated circular DNA molecules 1–4 (MarFaCM1-4) (Table 1). MarFaCM1 encodes a Rep protein and is most closely related to the Rep of marmot associated feces circular DNA molecule 4 and the Rep of CRESS DNA virus isolated from crucian carp tissue (MK012508) sharing 45% identity. MarFaCM3 encodes a CP that is most closely related to a putative CP cressdnavirus identified in a crucian tissue (MK012484) (Tisza et al., 2020). MarFaCM1 encodes a hypothetical viral-like protein and is most closely related to a hypothetical viral-like protein encoded by an Antarctic circular DNA molecule (MN32827) (Sommers et al., 2019). MarFaCM-2 encodes a putative CP that is most closely related that a gemycircularvirus isolate as3 (KF371630) (Sikorski et al., 2013). In some cases, those molecules may share high similarity in the intergenic region indicating they could be cognate molecules part of a possible multipartite viral entity (Kraberger et al., 2019; Male et al., 2016). MarFaCM1 and -2 were identified in the same sample, but shared no regions of similarity in the intergenic region. MarFaCM3 and -4 were also identified from the same individual sample and do not share any regions of similarity in the intergenic region.

All cressdnaviruses encode a distinct Rep that has conserved motifs, an origin of replication, and a CP. The Reps have conserved for HUH endonucleases and superfamily 3 helicases motifs (Rosario et al., 2012). We identified these motifs in the Reps of all the cressdnaviruses from this study and these are summarized in Table 1.

## 5. Concluding remarks

The identification of novel circular ssDNA viruses through high-throughput sequencing allows for a deeper understanding of the viruses associated with many species and ecosystems. Other than the anellovirus identified in this study, which likely infects the marmot, we are unable to say whether the rest of the virus identified in this study are associated with the diet, enteric microbes or environment. Nonetheless, we identified 26 cressdnaviruses and four circular DNA molecules (Fig. 1) associated with yellow-bellied marmots in this Colorado, USA population.

## CRediT authorship contribution statement

**Anthony Khalifeh:** Methodology, Validation, Data curation, Investigation, Writing - original draft, Writing - review & editing, Visualization. **Daniel T. Blumstein:** Conceptualization, Methodology, Investigation, Resources, Writing - review & editing, Funding acquisition. **Rafaela S. Fontenele:** Methodology, Validation, Investigation, Writing - review & editing. **Kara Schmidlin:** Methodology, Validation, Investigation, Writing - review & editing. **Cécile Richet:** Methodology, Validation, Investigation, Writing - review & editing. **Simona Kraberger:** Methodology, Validation, Investigation, Data curation, Writing - review & editing, Visualization, Supervision. **Arvind Varsani:** Conceptualization, Methodology, Validation, Investigation, Resources, Data curation, Writing - review & editing, Visualization, Supervision, Funding acquisition.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

We thank Dana Williams and the rest of 2018's Team Marmot for help collecting samples. DTB was supported by NSF-DEB 1557130 as well as by NSF-DBI 1226713 to the RMBL. The molecular work was supported by a Startup grant awarded to AV from Arizona State University and Barrett Thesis Funding awarded to AK from Barrett, The Honors College, Arizona State University.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.virol.2020.12.017.

## GenBank accession #s

MT044319–MT044326, MT181524–MT181546

## References

Águeda-Pinto, A., Kraberger, S., Lund, M.C., Gortázar, C., McFadden, G., Varsani, A., Esteves, P.J., 2020. Coinfections of novel polyomavirus, anelloviruses and a recombinant strain of myxoma virus-MYXV-tol identified in iberian hares. Viruses 12, 340.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Ao, Y., Li, X., Li, L., Xie, X., Jin, D., Yu, J., Lu, S., Duan, Z., 2017. Two novel bocaparvovirus species identified in wild Himalayan marmots. Sci. China Life Sci. 60, 1348–1356.

Armitage, K.B., 2014. Marmot Biology: Sociality, Individual Fitness, and Population Dynamics. Cambridge University Press, Cambridge.

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., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477.

Biagini, P., Bendinelli, M., Hino, S., Kakkola, L., Mankertz, A., Niel, C., Okamoto, H., Raidal, S., Teo, C., Todd, D., 2011. Family Anelloviridae, Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier Scientific Publ. Co, pp. 331–341.

Blanton, J.D., Palmer, D., Dyer, J., Rupprecht, C.E., 2011. Rabies surveillance in the United States during 2010. J. Am. Vet. Med. Assoc. 239, 773–783.

Blumstein, D.T., 2013. Yellow-bellied marmots: insights from an emergent view of sociality. Philos. Trans. R. Soc. Lond. B Biol. Sci. 368, 20120349.

Cibulski, S.P., Teixeira, T.F., de Sales Lima, F.E., do Santos, H.F., Franco, A.C., Roehe, P. M., 2014. A novel Anelloviridae species detected in Tadarida brasiliensis bats: first sequence of a chiropteran anellovirus. Genome Announc. 2 e01028-01014.

Crane, A., Goebel, M.E., Kraberger, S., Stone, A.C., Varsani, A., 2018. Novel anelloviruses identified in buccal swabs of Antarctic Fur seals. Virus Gene. 54, 719–723.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27, 1164–1165.

Dayaram, A., Goldstien, S., Arguello-Astorga, G.R., Zawar-Reza, P., Gomez, C., Harding, J.S., Varsani, A., 2015. Diverse small circular DNA viruses circulating amongst estuarine molluscs. Infect. Genet. Evol. 31, 284–295.

de Souza, W.M., Fumagalli, M.J., de Araujo, J., Sabino-Santos Jr., G., Maia, F.G.M., Romeiro, M.F., Modha, S., Nardi, M.S., Queiroz, L.H., Durigon, E.L., Nunes, M.R.T., Murcia, P.R., Figueiredo, L.T.M., 2018. Discovery of novel anelloviruses in small mammals expands the host range and diversity of the Anelloviridae. Virology 514, 9–17.

Edgar, R., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput.(2004). Nucleic Acids Res. 32, 1792–1797. PMID.

Fontenele, R.S., Lacorte, C., Lamas, N.S., Schmidlin, K., Varsani, A., Ribeiro, S.G., 2019. Single stranded DNA viruses associated with capybara faeces sampled in Brazil. Viruses 11, 710.

Frase, B.A., Hoffmann, R.S., 1980. Marmota flaviventris. Mamm. Species 1.

Freer, G., Maggi, F., Pifferi, M., Di Cicco, M.E., Peroni, D.G., Pistello, M., 2018. The virome and its major component, anellovirus, a convoluted system molding human immune defenses and possibly affecting the development of asthma and respiratory diseases in childhood. Front. Microbiol. 9, 686.

Gerlt, J.A., Bouvier, J.T., Davidson, D.B., Imker, H.J., Sadkhin, B., Slater, D.R., Whalen, K.L., 2015. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): a web tool for generating protein sequence similarity networks. Biochim. Biophys. Acta 1854, 1019–1037.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321.

Heissenberger, S., de Pinho, G.M., Martin, J.G.A., Blumstein, D.T., 2020. Age and location influence the costs of compensatory and accelerated growth in a hibernating mammal. Behav. Ecol. 31, 826–833.

Hrazdilova, K., Slaninkova, E., Brozova, K., Modry, D., Vodicka, R., Celer, V., 2016. New species of Torque Teno miniviruses infecting gorillas and chimpanzees. Virology 487, 207–214.
Kraberger, S., Arguello-Astorga, G.R., Greenfield, L.G., Galilee, C., Law, D., Martin, D.P., Varsani, A., 2015. Characterisation of a diverse range of circular replication-associated protein encoding DNA viruses recovered from a sewage treatment oxidation pond. Infect. Genet. Evol. 31, 73–86.
Kraberger, S., Cook, C.N., Schmidlin, K., Fontenele, R.S., Bautista, J., Smith, B., Varsani, A., 2019. Diverse single-stranded DNA viruses associated with honey bees (Apis mellifera). Infect. Genet. Evol. 71, 179–188.
Kraberger, S., Hofstetter, R.W., Potter, K.A., Farkas, K., Varsani, A., 2018a. Genomoviruses associated with mountain and western pine beetles. Virus Res. 256, 17–20.
Kraberger, S., Waits, K., Ivan, J., Newkirk, E., VandeWoude, S., Varsani, A., 2018b. Identification of circular single-stranded DNA viruses in faecal samples of Canada lynx (Lynx canadensis), moose (Alces alces) and snowshoe hare (Lepus americanus) inhabiting the Colorado San Juan Mountains. Infect. Genet. Evol. 64, 1–8.
Krupovic, M., Ghabrial, S.A., Jiang, D., Varsani, A., 2016. Genomoviridae: a new family of widespread single-stranded DNA viruses. Arch. Virol. 161, 2633–2643.
Krupovic, M., Varsani, A., Kazlauskas, D., Breitbart, M., Delwart, E., Rosario, K., Yutin, N., Wolf, Y.I., Harrach, B., Zerbini, F.M., Dolja, V.V., Kuhn, J.H., Koonin, E.V., 2020. Cressdnaviricota: a virus phylum unifying seven families of rep-encoding viruses with single-stranded, circular DNA genomes. J. Virol. 94.
Labonte, J.M., Suttle, C.A., 2013. Previously unknown and highly divergent ssDNA viruses populate the oceans. ISME J. 7, 2169–2177.
Liu, J., Wennier, S., Zhang, L., McFadden, G., 2011. M062 is a host range factor essential for myxoma virus pathogenesis and functions as an antagonist of host SAMD9 in human cells. J. Virol. 85, 3270–3282.
Maldonado-Chaparro, A.A., Blumstein, D.T., Armitage, K.B., Childs, D.Z., 2018. Transient LTRE analysis reveals the demographic and trait-mediated processes that buffer population growth. Ecol. Lett. 21, 1693–1703.
Male, M.F., Kraberger, S., Stainton, D., Kami, V., Varsani, A., 2016. Cycloviruses, gemycircularviruses and other novel replication-associated protein encoding circular viruses in Pacific flying fox (Pteropus tonganus) faeces. Infect. Genet. Evol. 39, 279–292.
McLean, D.M., Ladyman, S.R., Purvin-Good, K.W., 1968. Westward extension of Powassan virus prevalence. Can. Med. Assoc. J. 98, 946–949.
Muhire, B.M., Varsani, A., Martin, D.P., 2014. SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. PLoS One 9, e108277.
Nishiyama, S., Dutia, B.M., Sharp, C.P., 2015. Complete genome sequences of novel anelloviruses from laboratory rats. Genome Announc. 3 e01262-01214.
Nishiyama, S., Dutia, B.M., Stewart, J.P., Meredith, A.L., Shaw, D.J., Simmonds, P., Sharp, C.P., 2014. Identification of novel anelloviruses with broad diversity in UK rodents. J. Gen. Virol. 95, 1544–1553.
Orton, J.P., Morales, M., Fontenele, R.S., Schmidlin, K., Kraberger, S., Leavitt, D.J., Webster, T.H., Wilson, M.A., Kusumi, K., Dolby, G.A., Varsani, A., 2020. Virus discovery in desert tortoise fecal samples: novel circular single-stranded DNA viruses. Viruses 12, 143.
Ozgul, A., Childs, D.Z., Oli, M.K., Armitage, K.B., Blumstein, D.T., Olson, L.E., Tuljapurkar, S., Coulson, T., 2010. Coupled dynamics of body mass and population growth in response to environmental change. Nature 466, 482–485.
Price, M.N., Dehal, P.S., Arkin, A.P., 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS One 5, e9490.
Rosario, K., Duffy, S., Breitbart, M., 2012. A field guide to eukaryotic circular single-stranded DNA viruses: insights gained from metagenomics. Arch. Virol. 157, 1851–1871.
Rosario, K., Mettel, K.A., Benner, B.E., Johnson, R., Scott, C., Yusseff-Vanegas, S.Z., Baker, C.C., Cassill, D.L., Storer, C., Varsani, A., 2018. Virus discovery in all three major lineages of terrestrial arthropods highlights the diversity of single-stranded DNA viruses associated with invertebrates. PeerJ 6, e5761.
Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504.
Sikorski, A., Massaro, M., Kraberger, S., Young, L.M., Smalley, D., Martin, D.P., Varsani, A., 2013. Novel myco-like DNA viruses discovered in the faecal matter of various animals. Virus Res. 177, 209–216.
Sommers, P., Fontenele, R.S., Kringen, T., Kraberger, S., Porazinska, D.L., Darcy, J.L., Schmidt, S.K., Varsani, A., 2019. Single-stranded DNA viruses in antarctic cryoconite holes. Viruses 11, 1022.
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., 2016. Circular replication-associated protein encoding DNA viruses identified in the faecal matter of various animals in New Zealand. Infect. Genet. Evol. 43, 151–164.
Stover, B.C., Muller, K.F., 2010. TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. BMC Bioinf. 11, 7.
Tisza, M.J., Pastrana, D.V., Welch, N.L., Stewart, B., Peretti, A., Starrett, G.J., Pang, Y.-Y. S., Krishnamurthy, S.R., Pesavento, P.A., McDermott, D.H., 2020. Discovery of several thousand highly diverse circular DNA viruses. Elife 9.
Varsani, A., Kraberger, S., 2020. Create 17 genera and new 80 species (Anelloviridae). https://talk.ictvonline.org/files/proposals/animal_dna_viruses_and_retroviruses/m/animal_dna_ec_approved/10410/download.
Varsani, A., Krupovic, M., 2017. Sequence-based taxonomic framework for the classification of uncultured single-stranded DNA viruses of the family Genomoviridae. Virus Evol 3, vew037.
Wu, Z., Yang, L., Ren, X., He, G., Zhang, J., Yang, J., Qian, Z., Dong, J., Sun, L., Zhu, Y., Du, J., Yang, F., Zhang, S., Jin, Q., 2016. Deciphering the bat virome catalog to better understand the ecological diversity of bat viruses and the bat origin of emerging infectious diseases. ISME J. 10, 609–620.
Yu, X., Li, B., Fu, Y., Jiang, D., Ghabrial, S.A., Li, G., Peng, Y., Xie, J., Cheng, J., Huang, J., 2010. A geminivirus-related DNA mycovirus that confers hypovirulence to a plant pathogenic fungus. Proc. Natl. Acad. Sci. Unit. States Am. 107, 8387–8392.
Zhao, L., Rosario, K., Breitbart, M., Duffy, S., 2019. Eukaryotic circular rep-encoding single-stranded DNA (CRESS DNA) viruses: ubiquitous viruses with small genomes and a diverse host range. Adv. Virus Res. 103, 71–133.