



The virome of the white-winged vampire bat *Diaemus youngi* is rich in circular DNA viruses

André Alberto Witt^{1,2} · Raquel Silva Alves¹ · Juliana do Canto Olegário¹ · Laura Junqueira de Camargo¹ · Matheus Nunes Weber³ · Mariana Soares da Silva^{1,3} · Raíssa Canova¹ · Ana Cristina Sbaraini Mosena¹ · Samuel Paulo Cibulski⁴ · Ana Paula Mutterle Varela⁵ · Fabiana Quoos Mayer⁵ · Cláudio Wageck Canal¹ · Renata da Fontoura Budaszewski¹

Received: 25 July 2021 / Accepted: 17 March 2022 / Published online: 2 April 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

In the Neotropical region, the white-winged vampire bat (*Diaemus youngi*) is the rarest of the three species of vampire bats. This bat species feeds preferentially on bird blood, and there is limited information on the viruses infecting *D. youngi*. Hence, this study aimed to expand the knowledge about the viral diversity associated with *D. youngi* by sampling and pooling the lungs, liver, kidneys, heart, and intestines of all animals using high-throughput sequencing (HTS) on the Illumina MiSeq platform. A total of three complete and 10 nearly complete circular virus genomes were closely related to gemykrogvirus (*Genomoviridae* family), smacovirus (*Smacoviridae* family), and torque teno viruses (TTVs) (*Anelloviridae* family). In addition, three sequences of bat paramyxovirus were detected and found to be closely related to viruses reported in Pomona roundleaf bats and rodents. The present study provides a snapshot of the viral diversity associated with white-winged vampire bats and provides a baseline for comparison to viruses detected in future outbreaks.

Keywords Virome · Vampire bat · *Diaemus youngi* · High-throughput sequencing · South America

Edited by Simon D. Scott.

✉ Renata da Fontoura Budaszewski
virologyrocks@gmail.com

- ¹ Laboratório de Virologia, Faculdade de Veterinária, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Rio Grande do Sul, Brazil
- ² Secretaria Estadual de Agricultura, Pecuária e Desenvolvimento Rural (SEAPDR), Porto Alegre, Rio Grande do Sul, Brazil
- ³ Present Address: Laboratório de Microbiologia Molecular, Instituto de Ciências da Saúde, Universidade Feevale, Novo Hamburgo, Rio Grande do Sul, Brazil
- ⁴ Centro de Biotecnologia (Cbiotec), Universidade Federal da Paraíba, João Pessoa, Paraíba, Brazil
- ⁵ Laboratório de Biologia Molecular, Instituto de Pesquisas Veterinárias Desidério Finamor (IPVDF), Secretaria da Agricultura, Pecuária e Desenvolvimento Rural (SEAPDR), Eldorado Do Sul, Rio Grande do Sul, Brazil

Introduction

In recent decades, emerging viral diseases associated with chiropterans have caused the loss of human and animal lives worldwide, mainly in Africa, Asia, and Australia [1–3]. Chiropterans are excellent bioindicators of environmental quality [4]; however, anthropogenic changes, such as deforestation, intensive agriculture, and extensive livestock farming, have modified the environment, thus causing an imbalance in ecosystems and leading to close contact with animal species that may carry unknown putative pathogens. These factors lead to greater proximity and close contact among wild animals, livestock, and humans, thereby increasing the risk of virus spillover and emerging infectious disease events [5–7]. It is important to highlight that in the last three decades, important lethal human pathogens have emerged in bats, such as henipavirus [8, 9], ebolavirus [10], and coronaviruses, such as SARS-CoV [11], SARS-CoV-2 [12], and MERS-CoV [13]. Vampire bats represent important actors in the spread of the rabies virus and thus are a source of concern to health and animal production authorities [14, 15].

Recently, 248 new virus species have been reported from bats around the world [1]; however, few of them are known to have the potential to cause severe disease. Species of concern belong to RNA virus families, such as *Coronaviridae*, *Paramyxoviridae*, *Rhabdoviridae*, *Filoviridae*, and *Reoviridae* [6]. Studies conducted in Brazil have screened 16 bat species and reported 576 virus species within the *Adenoviridae*, *Astroviridae*, *Circoviridae*, *Coronaviridae*, *Hantaviridae*, *Papillomaviridae*, and *Parvoviridae* families according to data tracked by DBatVir (<http://www.mgc.ac.cn/DBatVir>).

In the Neotropical region, three species of vampire bats can be found: *Desmodus rotundus* (E. Geoffroy Saint-Hilaire, 1810), *Dyphilla ecaudata* (Spix, 1823), and *Diaemus youngi* (Jentink, 1893). Most of the research is focused on *D. rotundus* due to its wide distribution and high-density colonies [16–21]. Research on *D. caudata* is scarce [22–24], and to the best of our knowledge, the virus diversity and pathogen potential of *D. youngi* have not been reported, with the exception of herpesviruses detected in blood samples collected in French Guiana [25].

The recent and ongoing SARS-CoV-2 pandemic has put the role of bats as potential zoonotic virus reservoirs in the spotlight, thus leading to a race to identify potential biomes and hosts that could be harboring the next outbreak, including South American bats [26]. Tropical and subtropical regions are associated with various factors, such as evergreen forests, rich biomes, high mammal diversity and density, pasture areas and deforestation, and these factors make tropical and subtropical regions hotspots with a high risk of emergence of infectious diseases (EIDs) [27, 28].

Rapid evolution of high-throughput sequencing (HTS) technologies has led to the discovery of an ever-increasing genetic diversity of viral genomes from various sources, such as animals, plants, and environments. This approach is especially important for analyzing samples when little or no information is available on the virus diversity. The presence of *D. youngi* was recently reported for the first time in Rio Grande do Sul State, southern Brazil, by our research group [29]. Thus, the aim of this study was to take advantage of HTS and detect viruses in an “unbiased” manner by analyzing organ samples from *D. youngi*.

Materials and methods

Sample collection

In 2019, six *D. youngi* bats (one female and five males) were collected in the municipalities of Candelária, Restinga Seca, and São Miguel das Missões, which are located in Rio Grande do Sul State (Supplementary Fig. 1). The animals were captured with mist nets, euthanized with an overdose

of 0.5% thionembutal administered intraperitoneally (1 mL) and immediately necropsied. Samples of liver, kidney, lung, heart, and intestines were harvested, stored separately in cryovials and stored at $-20\text{ }^{\circ}\text{C}$ for up to two days, transported to the Veterinary Virology Laboratory and then stored at $-80\text{ }^{\circ}\text{C}$ until processing. The research authorization was obtained from the country's responsible environmental agency, the Chico Mendes Institute for Biodiversity Conservation (ICMBio), under authorization number 61537-1 (Supplementary material 1), to carry out the capture and collection of vampire bats occurring in several municipalities in Rio Grande do Sul.

All organs were macerated individually and diluted to 20% (w/v) in phosphate-buffered saline (PBS) (pH 7.2). One pool containing all the samples was assembled with 100 μL of each organ. PBS was added to the pool to a final volume of 10 mL, which was then centrifuged at low speed at $2000\times g$ for 30 min at $10\text{ }^{\circ}\text{C}$. The supernatant was filtered through a $0.45\text{ }\mu\text{m}$ filter to remove debris.

Viral metagenomics and HTS

The pool was ultracentrifuged on a 25% sucrose cushion at $150,000\times g$ for 3 h at $4\text{ }^{\circ}\text{C}$ in a Sorvall AH629 rotor. The pellet containing the viral particles was incubated for 1.5 h with DNase and RNase enzymes (Thermo Fisher Scientific, Waltham, MA, USA) [30]. Subsequently, viral RNA and DNA were isolated using TRI Reagent (Sigma Aldrich) and a standard phenol–chloroform protocol [31], respectively.

The viral DNA was enriched with the GenomePlex® Complete Whole Genome Amplification (WGA) Kit (Sigma-Aldrich, St. Louis, MO, USA), while the viral RNA was reverse-transcribed and enriched to dsDNA using the TransPlex® Complete Whole Transcriptome Amplification (WTA) Kit (Sigma-Aldrich, St. Louis, MO, USA), following the manufacturer's recommendations. The DNA products produced from these enrichment protocols were purified using the PCR Purification Combo Kit (Thermo Fisher Scientific). The quality and quantity of the DNA were assessed through spectrophotometry and fluorometry performed with a NanoDrop™ system (Thermo Fisher Scientific) and a Qubit™ system (Thermo Fisher Scientific), respectively, and the products were pooled in equimolar amounts to a final concentration of 0.2 ng of purified DNA. The library was further prepared using the Nextera XT DNA Library Preparation Kit and sequenced using an Illumina MiSeq System using an Illumina v2 reagent kit (2×150 paired-end reads).

Bioinformatic analysis

The quality of the generated reads was evaluated, adapters were trimmed using FASTQ Toolkit v. 2.2.5 (BaseSpace Labs), and the data were de novo assembled using

SPAdes Genome Assembler v. 3.9.0 [32]. Both tools were accessed on BaseSpace Sequence Hub (<https://basespace.illumina.com>). The assembled contigs were examined for similarities with known sequences through BLASTX using Blast2GO [33], and all relevant assemblies were confirmed by mapping reads to contigs using Geneious Prime software 2020.0.5 (<https://www.geneious.com>). Sequences with E -values $\leq 10^{-3}$ were classified as likely to have originated from eukaryotic viruses, bacteria, phages, or unknown sources based on the taxonomic origin of the sequence with the best E -value. Gene nucleotide and protein comparisons were performed with the BLASTN and BLASTP programs (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) to identify the sequences that were most closely related to the viral contigs of interest. Complete ORFs were predicted and annotated using Geneious Prime. Sequences were replaced by their reverse complements where necessary to maintain all sequences in the same strand. In the annotation of circular genomes, for comparative analyses, we phased all sequences to the same start position of a selected reference protein.

Reference and/or representative sequences of viruses belonging to the families *Anelloviridae*, *Genomoviridae*, *Smacoviridae*, and *Paramyxoviridae* were obtained from GenBank and aligned with the sequences identified in the present using ClustalW [34] using MEGA6 software [35]. Taxonomic analysis was conducted by considering the pairwise distance matrixes (of either identity or divergence) calculated using MEGA6 according to cutoff values determined by ICTV guidelines specific for each family. Phylogenetic trees were constructed using MEGA6. For each phylogenetic tree, the best model was also calculated using MEGA6.

Detection of paramyxovirus and coronavirus by RT-PCR

All organs of the six bats were screened individually for a 494 bp fragment of the L gene of paramyxoviruses using a well-established broadly reactive PCR protocol described previously [36] to amplify and sequence a particular fragment highly used for phylogenetic analysis. Additionally, the samples were screened individually for a 440 bp fragment of the pol gene of coronaviruses using a pan-coronavirus protocol successfully described previously [37] to detect any coronavirus-related sequences. Primer sequences are specified in Supplementary Table 1.

Results

Overview

The DNA library generated a total of 108,036 reads that were de novo-assembled into 1,280 contigs. The contigs were compared with the GenBank nonredundant protein database through a BLASTX search conducted with an E -value cutoff of 10^{-5} in Blast2GO [33]. Using this approach, the vast majority (1045/1280; 82%) of sequences could not be classified (“no blast hit”). Host genome and bacterial and phage genomes corresponded to 15% (197/1,280) of the contigs, while the exogenous eukaryotic virus-related sequences corresponded to 3% (38/1280) of the contigs (Fig. 1a).

The majority of eukaryotic virus sequences observed were circular DNA viruses (35/38, 92%), most of which belonged to the *Anelloviridae* family (26/38, 68%), while ten sequences were CRESS belonging to two viral families: *Genomoviridae* (7/38, 18%) and *Smacoviridae* (2/38, 5%).

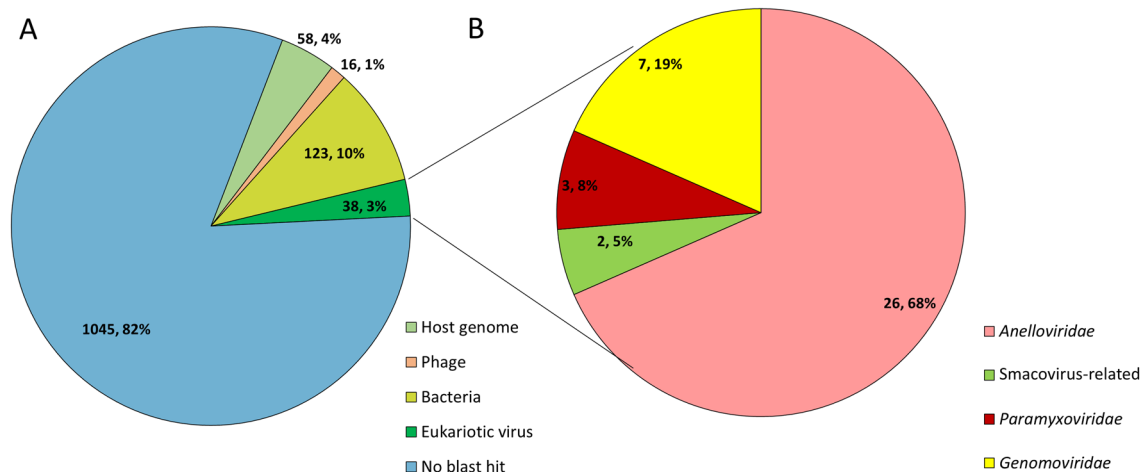


Fig. 1 Metagenomic graphic results presenting the generated sequence distribution. **a** All reads obtained from this analysis; and **b** eukaryotic virus representativeness

Three sequences (3/38, 8%) were related to single-stranded RNA (ssRNA) genomes belonging to the *Paramyxoviridae* family (Fig. 1b). Translated sequences similar to those of known or suspected eukaryotic viral proteins are briefly summarized in Table 1. The complete blastn/blastx results of virus-related contigs are presented in Supplementary Table 2. Information regarding the sequences obtained is described in the following sections.

Despite not being detected by HTS, a conventional RT-PCR “pan-coronavirus” protocol was performed to confirm its absence, and all animals were negative. A “pan-paramyxovirus” protocol was also performed, aiming for the phylogenetic analysis of a highly used L gene fragment, but despite positive results, repetitive attempts to sequence it by Sanger failed.

Anelloviridae

A total of 26 contigs closely related to *Anelloviridae* members were detected (Table 1), with eight classified as *Xitorquevirus*; moreover, three highly divergent contigs were proposed as belonging to a new genus herein named *Yodtorquevirus*. The contigs ranged from 250 to 2374 nt in length. It was possible to detect two complete genomes (Fig. 2a and b) and four nearly complete genomes that displayed three complete ORFs (ORF1, ORF2, and ORF3) (Fig. 2c and d). Five additional sequences displaying the complete ORF1 gene were also detected. Considering that the complete ORF1 nucleotide sequence is required for *Anelloviridae* member classification, the remaining 15 *Anelloviridae*-related contigs that displayed only partial ORF1 or only ORF2 were excluded from the analysis.

According to the most recent ICTV guidelines, the demarcation criteria for new virus species within family *Anelloviridae* were set to nucleotide identities of less than 69% (<http://www.ictvonline.org/virusTaxonomy.asp>) of

complete ORF1 nucleotide sequences, while new genera. Thus, the nucleotide pairwise distance was calculated using SDT 2.1 [38]. Our nucleotide pairwise identity comparison revealed that seven species were detected (Supplementary Fig. 2). Also, according to the ICTV guidelines, we took into consideration a phylogeny-based approach using the ORF1 amino acid sequences, and two distinct genera could be observed. The phylogenetic analysis was constructed based on the translated ORF1 of the 11 TTDyV, their most closely related sequences in GenBank and representative members of each genus (Fig. 3). The results indicated that the isolates of the present study were grouped into two main clades representing two viral genera: *Xitorquevirus* and a putative new genus proposed in the present study, that we named *Yodtorquevirus* according to ICTV suggestion of using the Phoenician alphabet. The genomes showed similar organization in ORF1, ORF2, and ORF3. The ORF1 sequences of the *Torque teno diaemus youngi virus* (TTDyV) ranged from 442 to 550 aa long, with typical arginine-rich regions at their N-termini (30% of the 70 first amino acids). A GC-rich region was found upstream of ORF2 for all analyzed sequences, while an extra GC region could be observed within ORF2 in the members of the putative new genus (Figs. 2b and d). Sequence analysis indicated that the 11 TTDyVs shared nucleotide sequence identity of their ORF1 ranging from 26.43 to 63.91%.

Three sequences were classified as two new species (TTDyV-6 and 7) within *Yodtorquevirus*, while one isolate (TTDyV-1) was closely related to TTVs previously described in bats from Brazil: one from a *Tadarida brasiliensis* (NC024908) collected in the same region as the present work [39] and one from *Desmodus rotundus* (MF541386) collected in São Paulo State (Southeast Brazil) in 2010 [20]. These three isolates belonged to the same species within the *Xitorquevirus* genus based on the pairwise distance (Supplementary table 2), and the seven remaining

Table 1 Summary of virus-related contigs detected in *Diaemus youngi* pool of tissues

Hits	Contigs lengths (nt)	Best blastx hits	Family	Genus ^a	Species ^a
26	236–2429	Torque teno virus	<i>Anelloviridae</i>	<i>Xitorquevirus</i>	Torque teno diaemus youngi virus 1 Torque teno diaemus youngi virus 2 Torque teno diaemus youngi virus 3 Torque teno diaemus youngi virus 4 Torque teno diaemus youngi virus 5
				<i>Yodtorquevirus^a</i>	Torque teno diaemus youngi virus 6 Torque teno diaemus youngi virus 7
7	330–2212	Gemycircularvirus	<i>Genomoviridae</i>	<i>Gemykrogvirus</i>	Giant panda-associated gemykrogvirus 1
2	478–2520	CRESS virus	<i>Smacoviridae</i>	Unclassified	Unclassified
3	332–464	Bat paramyxovirus, Mount Mabu virus 1	<i>Paramyxoviridae</i>	Unclassified	Unclassified

^aAs proposed in the present study

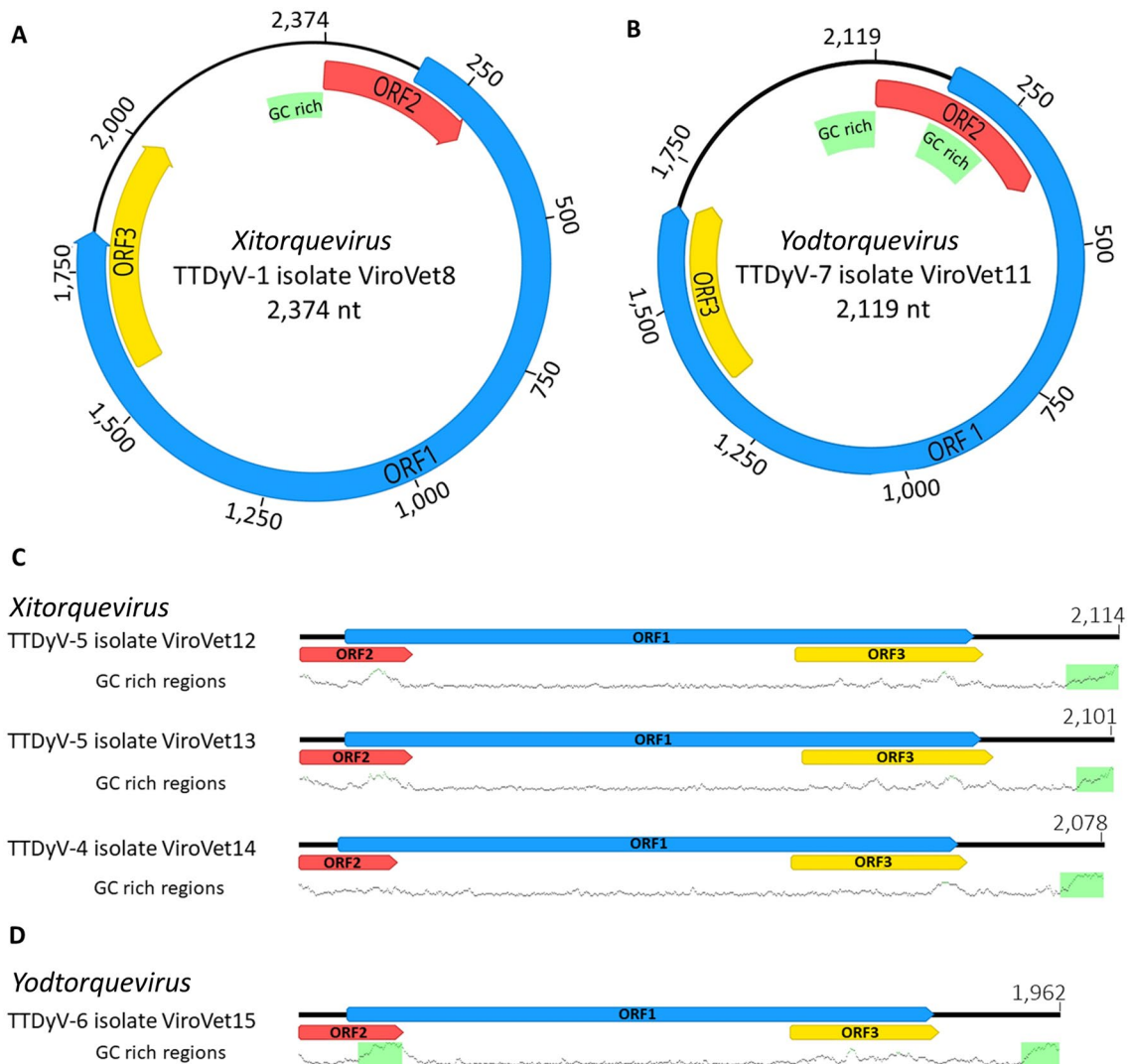


Fig. 2 **a** Genomic organization of TTDyV-1 (*Xitorquevirus*) and **b** TTDyV-7 (proposed *Yodtorquevirus* genus); and **c** linearized genomic organization of TTDyV- 4 and 5 (*Xitorquevirus*) and **d** TTDyV-6

(proposed *Yodtorquevirus* genus). The arrows represent the directions and reading frame of each putative ORF (ORF1-ORF3). A closed green box indicates the GC-rich regions

isolates detected in the present study comprised four new species (TTDyV-2, 3, 4, and 5) within this genus.

Genomoviridae

In the present study, seven contigs closely related to *Genomoviridae* members were detected (Table 1). The contigs ranged from 330 to 2212 nt in length. One complete genome of 2157 nt long was obtained (Fig. 4a). The genome organization included a spliced Rep protein ORF, a Cap ORF in the opposite orientation, and one putative ORF3 in the same orientation of Rep. The stem-loop structure was found between the 5' ends of the two main ORFs (Fig. 4a). The remaining contigs were of partial Rep/Cap and were analyzed only by means of blastn/blastx, as seen in Supplementary Table 2.

For taxonomic classification, the genome-wide pairwise identities and the phylogenetic tree of the translated Rep were analyzed. The phylogenetic analysis included our complete genome, the best matches based on Rep in GenBank and reference sequences considering the list from the family description from the ICTV (<http://talk.ictvonline.org/taxonomy/>), with a total of 47 sequences in the dataset.

Phasing of complete genomes was necessary to align the complete circular genomes, and the TATA box upstream of the stem loop was set as position 1. When phasing was not possible, the sequence was excluded from the analysis. The translated sequences were aligned with ClustalW [34] using MEGA6 software. Gap opening and extension penalties were set to 5 and 1, respectively, due to the high divergence of the sequences. A translated Rep maximum

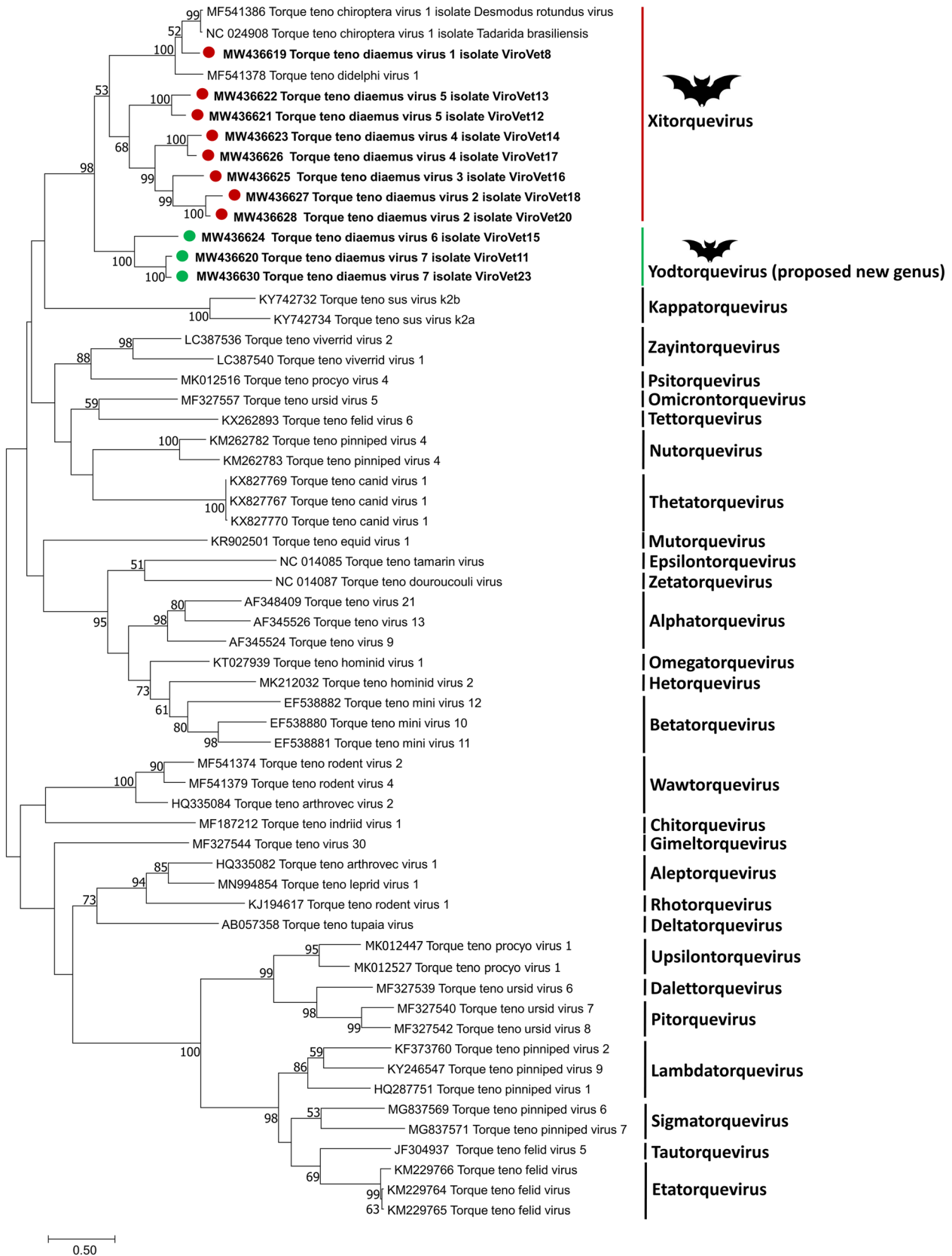


Fig. 3 Phylogenetic analysis performed based on the ORF1 nucleotide sequences of the 11 TTDyVs, their most closely related sequences in GenBank and representative members of each genus

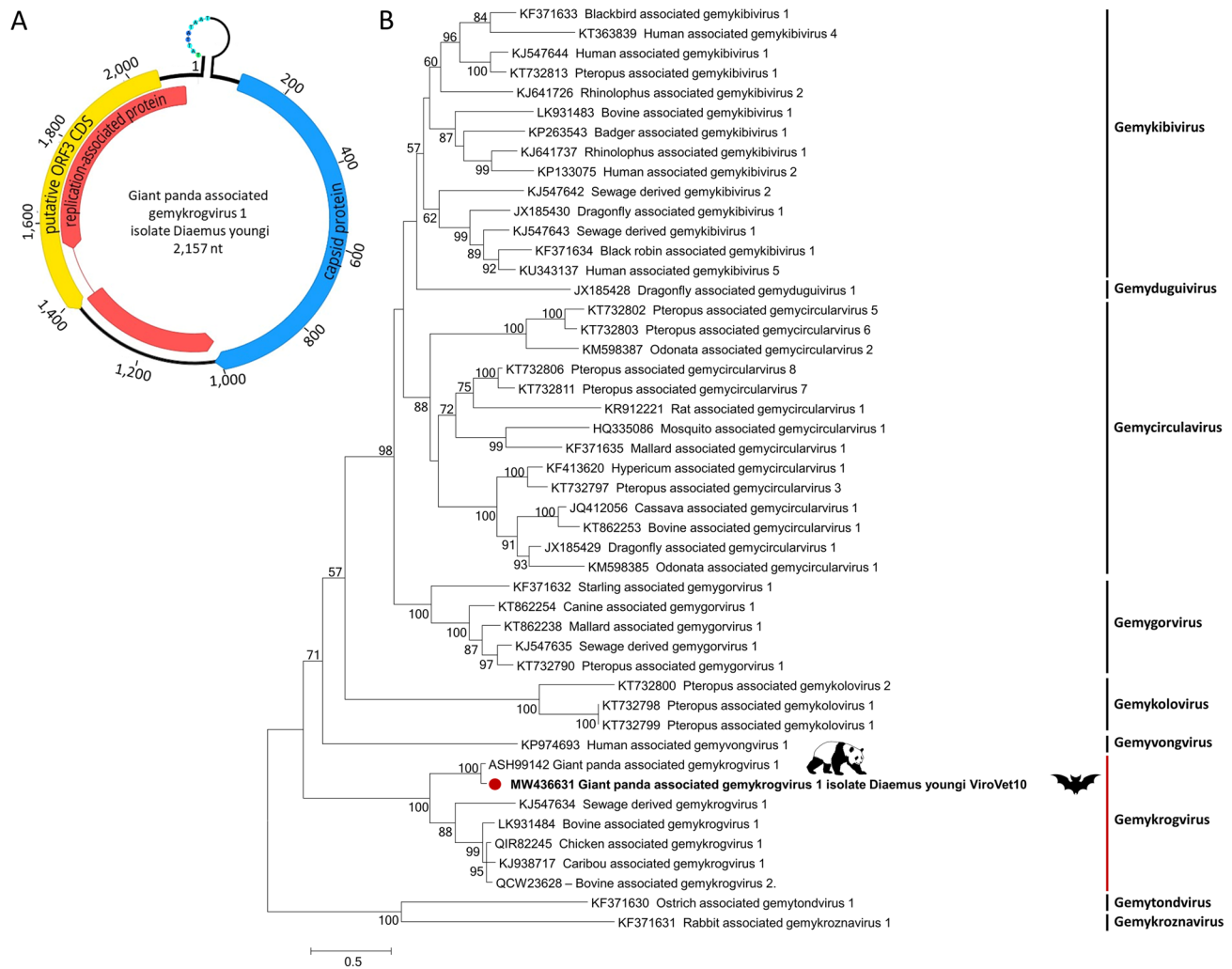


Fig. 4 **a** Genome map of *giant panda-associated gemykrogvirus 1* isolate *Diaemus youngi* (GiGemyV). Genes encoding the replication-initiation protein (Rep) and capsid protein (Cap) are shown with arrows. A putative ORF3 is also represented by an arrow. The position of the nonanucleotide (TAATATATT) at the potential stem-loop structure is also indicated; and **b** Rep amino acid phylogenetic tree of

Gemoviridae. The sequences were analyzed through the maximum likelihood method with the LG+G+I model. Analyses were conducted with 1000 bootstrap replicates. Bootstrap values higher than 50% are shown. The sequence detected in the present study is highlighted with a circle

likelihood tree was constructed (Fig. 4b), and genome-wide pairwise identities were calculated (one minus p distances of pairwise aligned sequences with pairwise deletion of gaps) for species classification (Supplementary Table 3). Following ICTV guidelines, 78% pairwise identity was set as a value for species demarcation. The genome was classified within the *Gemykrogvirus* genus, and it belonged to the same species as from a virus previously associated with a giant panda (GenBank accession number MF327559) [40] based on the 94.1% genome-wide pairwise identity (Supplementary table 3); for this reason, our genome was named *giant panda-associated gemykrogvirus 1* isolate *Diaemus youngi* ViroVet10 (GenBank accession number MW436631).

Paramyxoviridae

In the present study, three contigs closely related to the *Paramyxoviridae* family were detected (Table 1). The contig lengths were 332, 402, and 464 nt. The nodes were renamed ViroVet1, ViroVet2, and ViroVet3 for better reading of the study, and their mapping to a reference genome can be visualized on Fig. 5a. The blastx analysis (Supplementary Table 2) showed that ViroVet2 displayed 62.41% amino acid identity (query cover of 99%) with the L protein of an unclassified bat paramyxovirus (GenBank accession number AIF74184) detected in a Pomona roundleaf bat from China in 2012 (Wu et al., 2016); ViroVet1 had 64.55% identity (query cover

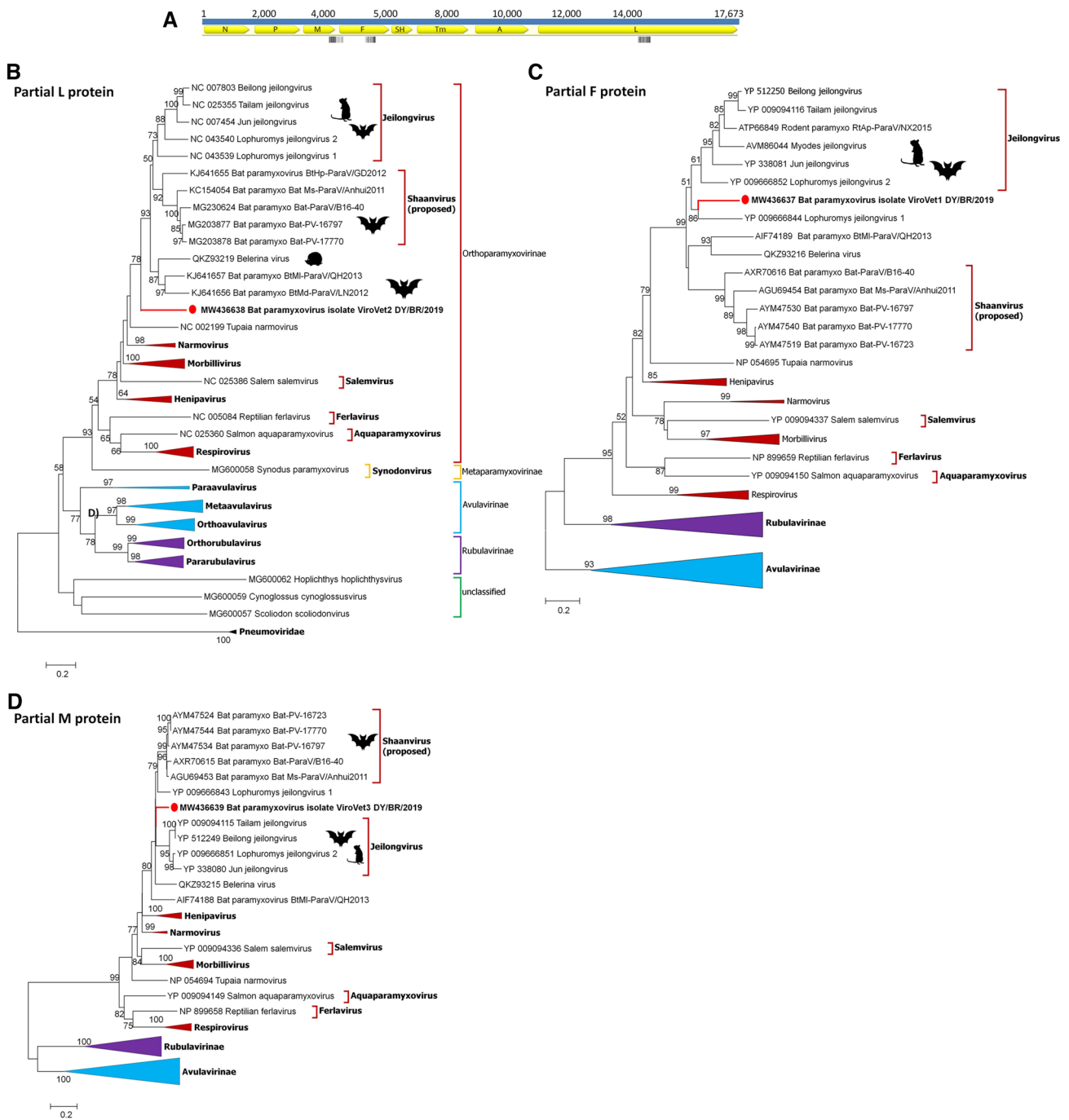


Fig. 5 **a** Mapping of bat paramyxovirus contigs ViroVet1, 2, and 3 in the complete genome of a reference bat paramyxovirus (MZ328288.1); **b** partial L; **c** partial F; and **d** partial M amino acid phylogenetic trees of *Paramyxoviridae* reference sequences and unclassified “bat paramyxovirus”. For the partial L tree, sequences of the *Pneumoviridae* family were included as outgroups. There were

86, 75, and 74 amino acid sequences, respectively, and 148 positions in the final dataset. The sequences were analyzed through the Maximum likelihood method with LG+G+I model. Analyses were conducted with 1000 bootstrap replicates. Bootstrap values higher than 50% are shown. The sequences detected in the present study were highlighted with circles

of 99%) with the F protein of *Lophumoris jeilongvirus 1* (formerly named Mount Mabu virus 1) detected in a Mozambique rodent in 2011 (GenBank accession number

YP_009666844) [41], while ViroVet3 had 81.16% identity (query cover of 44%) with the M protein of the same virus.

Based on the blastn results, the most similar sequences for each node were retrieved from the GenBank database along with reference sequences of the *Paramyxoviridae* family, with a final dataset of 86 sequences. Three datasets were analyzed separately. The nucleotide sequences were translated and aligned using ClustalW. A maximum likelihood phylogenetic tree was constructed for each partial protein.

The phylogenetic tree of the partial L protein (Fig. 5b) showed that ViroVet2 belongs to the *Orthoparmyxovirinae* subfamily but does not cluster close to any recognized genera. A cluster containing the newly recognized *Jeilongvirus*, the proposed *Shannvirus*, the newly described Balerina virus (unclassified) and unclassified bat paramyxoviruses is the most closely related to this L fragment. The phylogeny of the ViroVet1 partial F protein and ViroVet3 partial M protein (Fig. 5c and d, respectively) showed that our fragments were most related to *Jeilongvirus* members but still divergent. It was not possible to determine if the three fragments belonged to the same genome sequence or multiple genomes.

Smacoviridae

In the present study, two contigs closely related to *Smacoviridae* members were detected, one with 478 nt (partial Rep gene) and a full-length circular genome of 2465 nt, which

was named *Diaemus youngi-associated smacovirus-related virus* (DiaemusSV) (Fig. 6a). This genome encoded a Rep ORF and a Cap ORF in the same orientation, which was classified as the type V organization following the CRESS DNA virus classification scheme proposed by Rosario and collaborators [42]. A putative stem loop structure was located near the 3'-end of the Rep ORF with homology to the degenerate NANNNTTAC nonanucleotide sequence motif, which is also shared by smacoviruses [43].

The ICTV latest taxonomy mentions the existence of unidirectional smacoviruses, but they are not contemplated in the classification criteria. To the best of our knowledge, none unisense genome have yet been classified at the genus or species level. Thus, we conducted our analysis separately for each protein, and the virus was described as smacovirus-related.

A blast analysis of the complete genome showed that the most similar sequence was a unisense genome retrieved from a tracheal swab of a chicken in 2017 in the USA (GenBank accession number MN379594), and it had an identity of 88.28% in a query cover of 75%. In addition, 46 reference sequences were retrieved from GenBank, and 22 complete unisense genomes were included in the analysis. For the Rep phylogeny, amino acid sequences of *Banana bunchy top virus* (NP_604483) and *Cardamom bushy dwarf virus* (AHF47677) were added as outgroups representing the

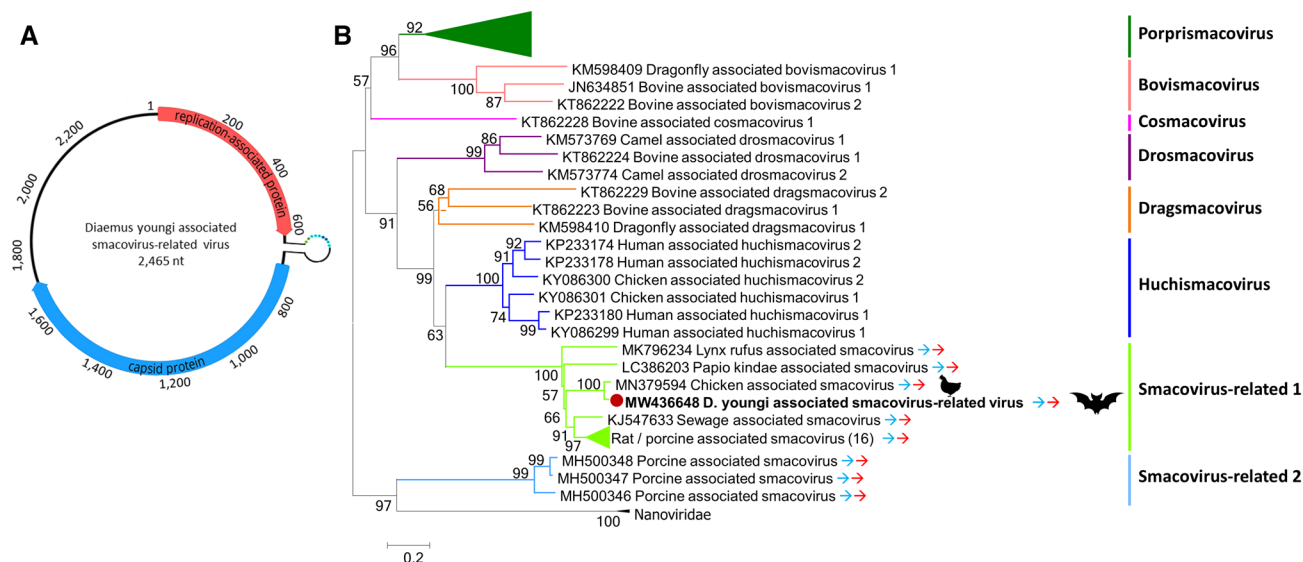


Fig. 6 **a** Genome map of *Diaemus youngi*-associated smacovirus-related virus. Genes encoding the replication-initiation protein (Rep) and capsid protein (Cap) are shown with arrows. The position of the nonanucleotide at the potential stem-loop structure is also indicated; and **b** Phylogenetic analysis of 69 translated Rep sequences of smacoviruses. The smacovirus described in the present study is highlighted with a circle. Two arrows indicate the genomes with unisense orientation. The evolutionary history was inferred by using the maximum likelihood method based on the LG model. A discrete gamma distri-

bution was used to model evolutionary rate differences among sites (5 categories (+G, parameter=3.4474)). Amino acid sequences of *Banana bunchy top virus* (NP604483) and *Cardamom bushy dwarf virus* (AHF47677) were used as outgroups representing the *Nanoviridae* family. All positions with less than 95% site coverage were eliminated. There were a total of 196 positions in the final dataset. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Evolutionary analyses were conducted in MEGA6

Nanoviridae family. Rep and Cap proteins were translated and aligned separately using ClustalW with gap opening and extension penalties of 5 and 1, respectively. Phylogenetic analyses were conducted using the maximum likelihood method within MEGA6.

The phylogenetic analysis of the Rep protein (Fig. 6b) showed that sequences derived from unisense genomes clustered in two separate groups, herein named smacovirus-related 1 and smacovirus-related 2, with high support values, one being highly divergent and closer to the outgroup. Our sequence, named isolate DiaemusSV, grouped within the smacovirus-related 1 cluster along with viruses associated with chicken, porcine, rodent, sewage, and other hosts (Fig. 8).

The Rep pairwise amino acid identity among the smacovirus-related 1 ranged from 58.4 to 100% (Supplementary Table 4), with the sequence chicken-associated smacovirus (GenBank accession number MN379594) being the closest to DiaemusSV, with 95.3% identity. The identity to established genera ranged from 15.2 to 37.1%, with *Porprisma-covirus* and *Huchismacovirus*, respectively.

Blast analysis of the Cap gene alone showed that its identity was higher with another chicken smacovirus isolated in the USA in 2017 (GenBank accession number MN379623.1), with 70.18% identity in a query cover of 98%. Interestingly, this sequence is ambisense. The blastn/blastx analysis of the partial Rep contig (Supplementary Table 2) showed that it was similar to a CRESS genome detected from a minnow in 2017 in the USA (GenBank accession number MH617713.1) by the same study group that identified the chicken smacoviruses described above. A phylogenetic tree of the translated Cap was constructed using the respective genes of the Rep dataset and excluding the outgroup sequences. Similar to the blast results, DiaemusSV grouped closely related to a chicken smacovirus that belongs to the *Porprisma-covirus* genus. The Cap phylogeny (Supplementary Fig. 5) showed great incongruency in relation to the Rep gene, suggesting frequent recombination events. A recombination analysis was not conducted here due to the need for alignment of complete genomes, which is not possible between viruses with opposite senses.

Discussion

The present study analyzed the virome of *D. youngi*, a vampire bat that feeds on bird blood and is widespread in Latin America, although it was only recently observed for the first time in Rio Grande do Sul State, Brazil. To the best of our knowledge, this is the largest and most up-to-date study analyzing virus diversity of the white-winged vampire bat. Recently, a study conducted in French Guiana described the diversity of herpesviruses in the same bat species [25],

although interestingly, this virus was not detected in the present work.

In the present study, we were able to detect 39 virus-related complete genomes or partial sequences, most of which were DNA circular viruses classified as anellovirus, genomovirus, and smacovirus-related and RNA genomes belonging to the *Paramyxoviridae* family. A possible explanation for the prevalence of circular ssDNA viruses might be the so-called ‘minimal lifestyle’ observed in some animal species and environments, as suggested previously [42]. ssDNA viruses, especially with circular genomes, seem to be widespread in some animal species and environments; torque teno viruses, for instance, are the most abundant component of the human virome [44].

Within the anelloviruses, we detected five new species within the previously proposed *Xitorquevirus* [20, 39], a genus not yet recognized by the taxonomic committee. Members of this group were first described from organs of a *Tadarida brasiliensis* bat also detected in Rio Grande do Sul State in 2013 [39]. Later, similar viruses were described in opossum (*Didelphis albiventris*) and bat species (*D. rotundus* and *Carollia perspicillata*) from the southeastern region of Brazil, and the *Sigma* genus was proposed [20]. In the present study, three anelloviruses were highly divergent and a new genus named *Yodtorquevirus* was proposed. Considering the distance-matrix values, we also proposed two different species: *Torque teno diaemus youngi virus 6* and 7. Our findings highlight the prevalence of torque teno viruses in the white-winged vampire bat virome.

In 2018, the circular replication-initiation protein encoding single-stranded (CRESS) DNA viruses associated with eukaryotic hosts were classified by ICTV into six families: *Circoviridae*, *Genomoviridae*, *Geminiviridae*, and *Nanoviridae* and the recently added *Bacilladnaviridae* and *Smacoviridae*. Currently consisting of nine genera and 73 species, the *Genomoviridae* family is found in a wide range of body fluids, tissues, organisms, and environments, and to date, there is no evidence of linked pathogenesis. Thus, genomoviridae viruses are usually named according to their source as “source-associated”, and this nomenclature will be used herein. Genomoviruses have an ambisense genome organization with two major inversely arranged ORFs encoding the replication-associated protein gene (Rep) and the capsid protein gene (Cap). A conserved stem-loop structure required for viral replication is located between the 5' ends of the two main ORFs.

To the best of our knowledge, this report presents the first detection of a genomovirus in white-winged bats and the first detection of *Gemykrogvirus* in bats. Members of this family have already been detected in insectivorous and frugivorous bat species: *Gemycirculavirus* has been detected in the intermediate roundleaf bat (*Hipposideros larvatus*) in China in 2016 (unpublished), in the Pacific flying bat

(*Pteropus tonganus*), in Tonga in 2015 [45], in the Eastern bent-wing bat (*Miniopterus fuliginosus*) in China in 2012 [46] and in European vespertilionid bats [47]. *Gemykibivirus* has been described in the greater horseshoe bat (*Rhinolophus ferrumequinum*) in China in 2013 [46] and in a Brazilian free-tailed bat (*Tadarida brasiliensis*) in Argentina in 2006.

The gemykrogvirus complete genome detected was surprisingly similar (94.1% genome-wide pairwise identity) to a virus previously isolated from the feces of a healthy captive giant panda in 2017 in China [40]. Other gemykrogviruses related to our sequence were detected in sewage, bovines, poultry, and reindeers. However, the virus was likely not acquired directly from the bat food source (chicken blood) but rather from a common alimentary source found in the environment because genomoviruses are found in a wide variety of organisms and environments. To date, only fungi have been found to be definitive hosts of a genomovirus of the *Gemycirculavirus* genus [48]. Considering that our sample was a pool of organs, including intestines, we could only assume whether the source was animal tissue or if it was present only in the fecal matter.

Smacoviruses were first detected in wild-living chimpanzee stool [49] and later described in the feces of healthy and diarrheic pigs [50], cattle [51], poultry [52], humans [53], and insects. Although these viruses are thought to infect eukaryotes, their actual host has not yet been confirmed. To date, only the fecal archaeon *Candidatus Methanomassiliicoccus intestinalis* has been shown to be a candidate host [54]. However, it remains to be established whether smacoviruses infect mammalian cells and cause disease.

The *Smacoviridae* family is currently classified into six genera. Smacoviruses have 2.3–2.9 kb genomes that contain two major ORFs encoding Rep and Cap proteins, typically in an ambisense organization, although unisense genomes have been reported [55]. The genetic diversity observed within smacoviruses might be due to high mutation rates and intra-familial recombination events in their genomes [56]. To the best of our knowledge, unisense genomes have not been classified or further characterized thus far, which is probably due to genome-wide alignment limitations required for analysis. Thus, we sought to analyze translated Reps and Caps separately. The phylogenetic analysis of the Rep protein showed that sequences derived from unisense genomes clustered in two separate groups with high support values, and one was highly divergent and closer to the outgroup. Considering the lack of guidelines for unisense genomes taxonomy, we classified unisense sequences, including our genome, as smacovirus-related 1 and smacovirus-related 2. The presence of smacovirus unisense genomes should be taken into consideration in further classification criteria.

Bats have been recently highly recognized as a potential threat for virus spillover due to the coronavirus pandemic

caused by SARS-CoV-2 because the most closely related viruses to SARS-CoV-2 were of bat origin [12]. Notably, bats also host major mammalian paramyxoviruses, including Nipah and Hendra viruses, members of the *Henipavirus* genus that are responsible for highly transmissible and lethal respiratory outbreaks. Recently, three new genera of orthoparamyxoviruses have been defined by the ICTV (<https://talk.ictvonline.org/taxonomy/>), i.e., *Jeilongvirus*, *Narmovirus*, and *Salemvirus*, while hundreds of unclassified “bat paramyxoviruses” are being uploaded to the GenBank database and increasingly reported worldwide [57–59]. A cluster of bat paramyxoviruses has been proposed as a new genus named *Shaanvirus* [46], and together with the genus *Jeilongvirus*, they encompass a large number of bat- and rodent-associated viruses. In the present study, we detected three partial sequences of the L, M, and F proteins of a bat paramyxovirus. The L gene encodes the polymerase protein, and it is the most conserved gene of paramyxoviruses and thus is commonly used for classification. The phylogenetic tree of the partial L protein showed that it belongs to a highly divergent paramyxovirus within the *Orthoparamyxovirinae* family, which could not be further classified. The M and F genes are known to be more variable, which explains the incongruency of the phylogenies between the three fragments. Novel attempts to amplify and sequence a larger fragment of the L gene and remaining genes are necessary to elucidate the classification of this highly divergent paramyxovirus.

Conclusion

A large number of ssDNA circular viruses with high diversity were described, and a new anellovirus genus, *Yodtorquevirus*, and seven new torque teno species were proposed. Additionally, genomoviruses and smacovirus-related sequences were described for the first time in this bat species. A highly divergent paramyxovirus was detected; however, further studies are necessary. Currently, the zoonotic risk for these viruses is low since their genomes do not display close phylogenetic relationships to viruses detected in humans; however, continuous surveillance is of paramount importance for prevention. Our study provides the first overview of the virome of the white-winged vampire bat and significantly increases the diversity of viruses.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11262-022-01897-6>.

Acknowledgements Financial support was provided by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Finance code 001), Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Propesq/UFRGS. The authors would

like to express their gratitude to the official Animal Sanitary Defense service of the State Secretary of Agriculture of Rio Grande do Sul State for generously providing samples for analysis and to the graduate and postgraduate students of the Laboratório de Virologia for their collaboration in this work.

Declarations

Conflict of interest The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Consent for publication All contributing authors have read and approved of the final version of the manuscript.

Ethical approval The project was authorized by ICMBio (Chico Mendes Institute for Biodiversity Conservation) under license number #61537-1. Animal collection was supported by the official service of Animal Sanitary Defense of the State Secretary of Agriculture of Rio Grande do Sul State. All applicable international, national and/or institutional guidelines for the care and use of animals were followed.

References

- Wang L-F, Anderson DE (2019) Viruses in bats and potential spillover to animals and humans. *Curr Opin Virol* 34:79–89. <https://doi.org/10.1016/j.coviro.2018.12.007>
- Smith I, Wang L-F (2013) Bats and their virome: an important source of emerging viruses capable of infecting humans. *Curr Opin Virol* 3:84–91. <https://doi.org/10.1016/j.coviro.2012.11.006>
- Brook CE, Dobson AP (2015) Bats as “special” reservoirs for emerging zoonotic pathogens. *Trends Microbiol* 23:172–180. <https://doi.org/10.1016/j.tim.2014.12.004>
- Jones G, Jacobs DS, Kunz TH et al (2009) Carpe noctem: the importance of bats as bioindicators. *Endang Species Res* 8:93–115. <https://doi.org/10.3354/esr00182>
- Jones BA, Grace D, Kock R et al (2013) Zoonosis emergence linked to agricultural intensification and environmental change. *PNAS* 110:8399–8404. <https://doi.org/10.1073/pnas.1208059110>
- Hayman DTS, Bowen RA, Cryan PM et al (2013) Ecology of zoonotic infectious diseases in bats: current knowledge and future directions. *Zoonoses Public Health* 60:2–21. <https://doi.org/10.1111/zph.12000>
- Dhaka P, Vijay D (2018) Role of bats in the emergence of viral zoonoses. *J Vet Anim Sci* 49(2):1–9
- Murray K, Selleck P, Hooper P et al (1995) A morbillivirus that caused fatal disease in horses and humans. *Science* 268:94–97. <https://doi.org/10.1126/science.7701348>
- Chua KB (2003) Nipah virus outbreak in Malaysia. *J Clin Virol* 26:265–275. [https://doi.org/10.1016/S1386-6532\(02\)00268-8](https://doi.org/10.1016/S1386-6532(02)00268-8)
- Leroy EM, Kumulungui B, Pourrut X et al (2005) Fruit bats as reservoirs of Ebola virus. *Nature* 438:575–576. <https://doi.org/10.1038/438575a>
- Kuiken T, Fouchier RAM, Schutten Met al (2003) Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome. *Lancet* 362:263–270. [https://doi.org/10.1016/S0140-6736\(03\)13967-0](https://doi.org/10.1016/S0140-6736(03)13967-0)
- Zhou P, Yang X-L, Wang X-G et al (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579:270–273. <https://doi.org/10.1038/s41586-020-2012-7>
- Zaki AM, van Boheemen S, Bestebroer TM et al (2012) Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med*. <https://doi.org/10.1056/NEJMoa1211721>
- Zarza H, Martínez-Meyer E, Suzán G et al (2017) Geographic distribution of *Desmodus rotundus* in Mexico under current and future climate change scenarios: Implications for bovine paralytic rabies infection. *Vet Méx OA*. <https://doi.org/10.21753/vmoa.4.3.390>
- Lee DN, Papeş M, van den Bussche RA (2012) Present and potential future distribution of common vampire bats in the Americas and the associated risk to cattle. *PLoS ONE* 7:e42466. <https://doi.org/10.1371/journal.pone.0042466>
- Quan P-L, Firth C, Conte JM et al (2013) Bats are a major natural reservoir for hepaciviruses and pegiviruses. *PNAS* 110:8194–8199. <https://doi.org/10.1073/pnas.1303037110>
- Lima FES, Cibulski SP, Elesbao F et al (2013) First detection of adenovirus in the vampire bat (*Desmodus rotundus*) in Brazil. *Virus Genes* 47:378–381. <https://doi.org/10.1007/s11262-013-0947-6>
- Drexler JF, Corman VM, Müller MA et al (2012) Bats host major mammalian paramyxoviruses. *Nat Commun* 3:1–3
- de Souza WM, Romeiro MF, Fumagalli MJ et al (2017) Chapparraviruses occur in at least three vertebrate classes and have a broad biogeographic distribution. *J Gen Virol* 98:225
- de Souza WM, Fumagalli MJ, de Araujo J et al (2018) Discovery of novel anelloviruses in small mammals expands the host range and diversity of the Anelloviridae. *Virology* 514:9–17. <https://doi.org/10.1016/j.virol.2017.11.001>
- Brandão PE, Scheffer K, Villarreal LY et al (2008) A coronavirus detected in the vampire bat *Desmodus rotundus*. *Braz J Infect Dis* 12:466–468. <https://doi.org/10.1590/s1413-867000800600003>
- Escalera-Zamudio M, Rojas-Anaya E, Kolokotronis S-O et al (2016) Bats, Primates, and the Evolutionary Origins and Diversification of Mammalian Gammaherpesviruses. *MBio*. <https://doi.org/10.1128/mBio.01425-16>
- Castilho JG, Carnieli P, Oliveira RN et al (2010) A comparative study of rabies virus isolates from hematophagous bats in Brazil. *J Wildl Dis* 46:1335–1339. <https://doi.org/10.7589/0090-3558-46.4.1335>
- de Araujo J, Thomazelli LM, Henriques DA et al (2012) Detection of hantavirus in bats from remaining rain forest in São Paulo. *Brazil BMC Res Notes* 5:690. <https://doi.org/10.1186/1756-0500-5-690>
- James S, Donato D, de Thoisy B et al (2020) Novel herpesviruses in neotropical bats and their relationship with other members of the Herpesviridae family. *Infect Genet Evol* 84:104367. <https://doi.org/10.1016/j.meegid.2020.104367>
- Cibulski S, de Lima FES, Roehe PM (2020) Coronaviruses in Brazilian bats: a matter of concern? *PLoS Negl Trop Dis* 14:e0008820. <https://doi.org/10.1371/journal.pntd.0008820>
- Olival KJ, Hosseini PR, Zambrana-Torrel C et al (2017) Host and viral traits predict zoonotic spillover from mammals. *Nature* 546:646–650. <https://doi.org/10.1038/nature22975>
- Morse SS, Mazet JAK, Woolhouse M et al (2012) Prediction and prevention of the next pandemic zoonosis. *The Lancet* 380:1956–1965. [https://doi.org/10.1016/S0140-6736\(12\)61684-5](https://doi.org/10.1016/S0140-6736(12)61684-5)
- Witt A, Mendes F, Grillo H et al (2019) Primeiro registro de *Diaemus youngi* (Jentink, 1893) no estado do Rio Grande do Sul, Brasil. *Notas Mam Sud* 1:1–7. <https://doi.org/10.31687/saremnms.19.0.15>
- Thurber RV, Haynes M, Breitbart M et al (2009) Laboratory procedures to generate viral metagenomes. *Nat Protoc* 4:470–483. <https://doi.org/10.1038/nprot.2009.10>

31. Sambrook J, & Russell, DW (2006) Purification of nucleic acids by extraction with phenol: chloroform. *CSH Proto* 2006(1):pdb.prot4455. <https://doi.org/10.1101/pdb.prot4455>
32. Bankevich A, Nurk S, Antipov D et al (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 14:455–477. <https://doi.org/10.1089/cmb.2012.0021>
33. Götz S, García-Gómez JM, Terol J et al (2008) High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res* 36:3420–3435. <https://doi.org/10.1093/nar/gkn176>
34. Larkin MA, Blackshields G, Brown NP et al (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948. <https://doi.org/10.1093/bioinformatics/btm404>
35. Tamura K, Stecher G, Peterson D et al (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30:2725–2729. <https://doi.org/10.1093/molbev/mst197>
36. Tong S, Chern S-WW, Li Y et al (2008) Sensitive and broadly reactive reverse transcription-PCR assays to detect novel paramyxoviruses. *J Clin Microbiol* 46:2652–2658. <https://doi.org/10.1128/JCM.00192-08>
37. Woo PCY, Lau SKP, Li KSM et al (2006) Molecular diversity of coronaviruses in bats. *Virology* 351:180–187. <https://doi.org/10.1016/j.virol.2006.02.041>
38. Muhire BM, Varsani A, & Martin DP (2014) SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. *PloS one* 9(9):e108277
39. Cibulski SP, Teixeira TF, Sales Lima FE, de, et al (2014) A novel Anelloviridae species detected in Tadarida Brasiliensis bats: first sequence of a chiropteran Anellovirus. *Genome Announc* <https://doi.org/10.1128/genomeA.01028-14>
40. Zhang W, Yang S, Shan T et al (2017) Virome comparisons in wild-diseased and healthy captive giant pandas. *Microbiome* 5:90. <https://doi.org/10.1186/s40168-017-0308-0>
41. Vanmechelen B, Vergote V, Merino M et al. (2020) Common occurrence of Belerina Virus, a novel paramyxovirus found in Belgian hedgehogs. *Proceedings* 50:40. Doi:<https://doi.org/10.3390/proceedings2020050040>
42. 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. <https://doi.org/10.1007/s00705-012-1391-y>
43. Tochetto C, Muterle Varela AP, Alves Lima de D et al (2020) Viral DNA genomes in sera of farrowing sows with or without stillbirths. *PLoS ONE* 15:e0230714. <https://doi.org/10.1371/journal.pone.0230714>
44. de Vlaminck I, Khush KK, Strehl C et al (2013) Temporal response of the human virome to immunosuppression and antiviral therapy. *Cell* 155:1178–1187. <https://doi.org/10.1016/j.cell.2013.10.034>
45. Male MF, Kraberger S, Stinton D et al (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. <https://doi.org/10.1016/j.meegid.2016.02.009>
46. Wu Z, Yang L, Ren X et al (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. <https://doi.org/10.1038/ismej.2015.138>
47. Kemenesi G, Kurucz K, Zana B et al (2018) Diverse replication-associated protein encoding circular DNA viruses in guano samples of Central-Eastern European bats. *Arch Virol* 163:671–678. <https://doi.org/10.1007/s00705-017-3678-5>
48. Yu X, Li B, Fu Y et al (2010) A geminivirus-related DNA mycovirus that confers hypovirulence to a plant pathogenic fungus. *PNAS* 107:8387–8392. <https://doi.org/10.1073/pnas.0913535107>
49. Blinkova O, Victoria J, Li Y et al (2010) Novel circular DNA viruses in stool samples of wild-living chimpanzees. *J Gen Virol* 91:74
50. Cheung AK, Ng TFF, Lager KM et al (2015) Identification of several clades of novel single-stranded circular DNA viruses with conserved stem-loop structures in pig feces. *Arch Virol* 160:353–358. <https://doi.org/10.1007/s00705-014-2234-9>
51. Kim HK, Park SJ, van Nguyen G et al (2012) Identification of a novel single-stranded, circular DNA virus from bovine stool. *J Gen Virol* 93:635–639. <https://doi.org/10.1099/vir.0.037838-0>
52. Lima DA, Cibulski SP, Finkler F et al (2017) Faecal virome of healthy chickens reveals a large diversity of the eukaryote viral community, including novel circular ssDNA viruses. *J Gen Virol* 98:690–703. <https://doi.org/10.1099/jgv.0.000711>
53. Ng TFF, Zhang W, Sachsenröder J et al (2015) A diverse group of small circular ssDNA viral genomes in human and non-human primate stools. *Virus Evol* 1:vev017. <https://doi.org/10.1093/ve/vev017>
54. Díez-Villaseñor C, Rodríguez-Valera F (2019) CRISPR analysis suggests that small circular single-stranded DNA smacoviruses infect Archaea instead of humans. *Nat Commun* 10:294. <https://doi.org/10.1038/s41467-018-08167-w>
55. Anindita PD, Sasaki M, Gonzalez G et al (2019) Discovery and genetic characterization of diverse smacoviruses in Zambian non-human primates. *Sci Rep* 9:5045. <https://doi.org/10.1038/s41598-019-41358-z>
56. 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:vev037. <https://doi.org/10.1093/ve/vev037>
57. Noh JY, Jeong DG, Yoon S-W et al (2018) Isolation and characterization of novel bat paramyxovirus B16–40 potentially belonging to the proposed genus Shaanvirus. *Sci Rep* 8:12533. <https://doi.org/10.1038/s41598-018-30319-7>
58. Jang SS, Noh JY, van Lo T et al (2020) The epidemiological characteristics of the Korean Bat Paramyxovirus between 2016 and 2019. *Microorganisms* 8:844. <https://doi.org/10.3390/microorganisms8060844>
59. Cibulski SP, Sales Lima FE, de, Teixeira TF, et al (2020) Detection of multiple viruses in oropharyngeal samples from Brazilian free-tailed bats (*Tadarida brasiliensis*) using viral metagenomics. *Arch Virol*. <https://doi.org/10.1007/s00705-020-04825-x>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.