



# Viral diversity in oral cavity from *Sapajus nigritus* by metagenomic analyses

Raissa Nunes dos Santos<sup>1,2</sup> · Fabricio Souza Campos<sup>2,3</sup> · Fernando Finoketti<sup>1,2</sup> · Anne Caroline dos Santos<sup>1,2</sup> · Aline Alves Scarpellini Campos<sup>1,2,3</sup> · Paulo Guilherme Carniel Wagner<sup>2,4</sup> · Paulo Michel Roehle<sup>1,2</sup> · Helena Beatriz de Carvalho Ruthner Batista<sup>2,5</sup> · Ana Claudia Franco<sup>1,2</sup>

Received: 20 January 2020 / Accepted: 25 July 2020 / Published online: 11 August 2020

© Sociedade Brasileira de Microbiologia 2020

## Abstract

*Sapajus nigritus* are non-human primates which are widespread in South America. They are omnivores and live in troops of up to 40 individuals. The oral cavity is one of the main entry routes for microorganisms, including viruses. Our study proposed the identification of viral sequences from oral swabs collected in a group of capuchin monkeys ( $n = 5$ ) living in a public park in a fragment of Mata Atlantica in South Brazil. Samples were submitted to nucleic acid extraction and enrichment, which was followed by the construction of libraries. After high-throughput sequencing and contig assembly, we used a pipeline to identify 11 viral families, which are *Herpesviridae*, *Parvoviridae*, *Papillomaviridae*, *Polyomaviridae*, *Caulimoviridae*, *Iridoviridae*, *Astroviridae*, *Poxviridae*, and *Baculoviridae*, in addition to two complete viral genomes of *Anelloviridae* and *Genomoviridae*. Some of these viruses were closely related to known viruses, while other fragments are more distantly related, with 50% of identity or less to the currently available virus sequences in databases. In addition to host-related viruses, insect and small vertebrate-related viruses were also found, as well as plant-related viruses, bringing insights about their diet. In conclusion, this viral metagenomic analysis reveals, for the first time, the profile of viruses in the oral cavity of wild, free ranging capuchin monkeys.

**Keywords** Oral virome · *Genomoviridae* · Cebidae · Wild fauna · *Anelloviridae*

Responsible Editor: Flavio Guimaraes Fonseca.

✉ Raissa Nunes dos Santos  
engraissanunes@gmail.com

<sup>1</sup> Laboratório de Virologia – Departamento de Microbiologia, Imunologia e Parasitologia, Universidade Federal do Rio Grande do Sul, Rua Sarmiento Leite, 500, Bairro Farroupilha, Porto Alegre, Rio Grande do Sul 90050-170, Brazil

<sup>2</sup> Centro Estadual de Vigilância em Saúde, Secretaria Estadual da Saúde do Rio Grande do Sul, Rua Domingos Crescêncio, 132, Bairro Santana, Porto Alegre, Rio Grande do Sul 90650-090, Brazil

<sup>3</sup> Universidade Federal de Tocantins, Rua Badejós, lote 7, Zona Rural, Gurupi, Tocantins 77402-970, Brazil

<sup>4</sup> IBAMA – Instituto Brasileiro de Meio Ambiente e dos Recursos Naturais Renováveis, Rua Miguel Teixeira, 216, Bairro Cidade Baixa, Porto Alegre, Rio Grande do Sul 90050-250, Brazil

<sup>5</sup> Instituto Pasteur – Secretaria da Saúde do Governo do estado de São Paulo, Av. Paulista, 393, Bairro Cerqueira César, São Paulo 01311-000, Brazil

## Introduction

Capuchin monkeys belong to the *Cebidae* family, which includes the genus *Sapajus* and comprises 29 species. These animals are small- to medium-sized monkeys (0.5–5.0 kg) [1] and their longevity in captivity is about 40 years. Their status is common and abundant, but declining due to clearance and fragmentation of forests. These animals are considered by the Red List Category and Criteria, according to the IUCN (International Union for Conservation of Nature) [2], near-threatened animals. Black-horned capuchins (*Sapajus nigritus*) are widespread in South America, including Colombia, Venezuela, Guyana, Paraguay, Brazil, and Argentina. The rapid growing and expansion of human populations drive the anthropogenic pressures (agriculture, hunting, and tourism), leading to close interactions between these animals and local people and increasing the risk of injuries by their impulsive behavior [3].

The oral cavity is a portal of entry for several microorganisms, and it contains the gingiva, tongue, palate, the surface of teeth,

and saliva. The microorganisms which colonize this environment suffer different pressures (including abiotic factors, osmotic conditions, and pH variation), which are related to the composition of food intake [4]. Metagenomic approaches, used in order to characterize viruses in human oral cavities, reveal that oral assemblages are largely composed by bacteriophages [5, 6]. In addition to these, eukaryotic viruses can also be found, as they may replicate in or close to the epithelium of the oral cavity and may be excreted by the saliva.

The periodic discovery of new viruses suggests that we underestimate the total number and diversity of viruses existing in nature [7]. On the other hand, the growing improvement of virus detection contributes to the discovery of a microbiology world yet unexplored. High-throughput sequencing (HTS) has facilitated the increase on the knowledge about viruses, being a robust and potential tool to better understand the circulation and potential spread of viruses from wild animals [8].

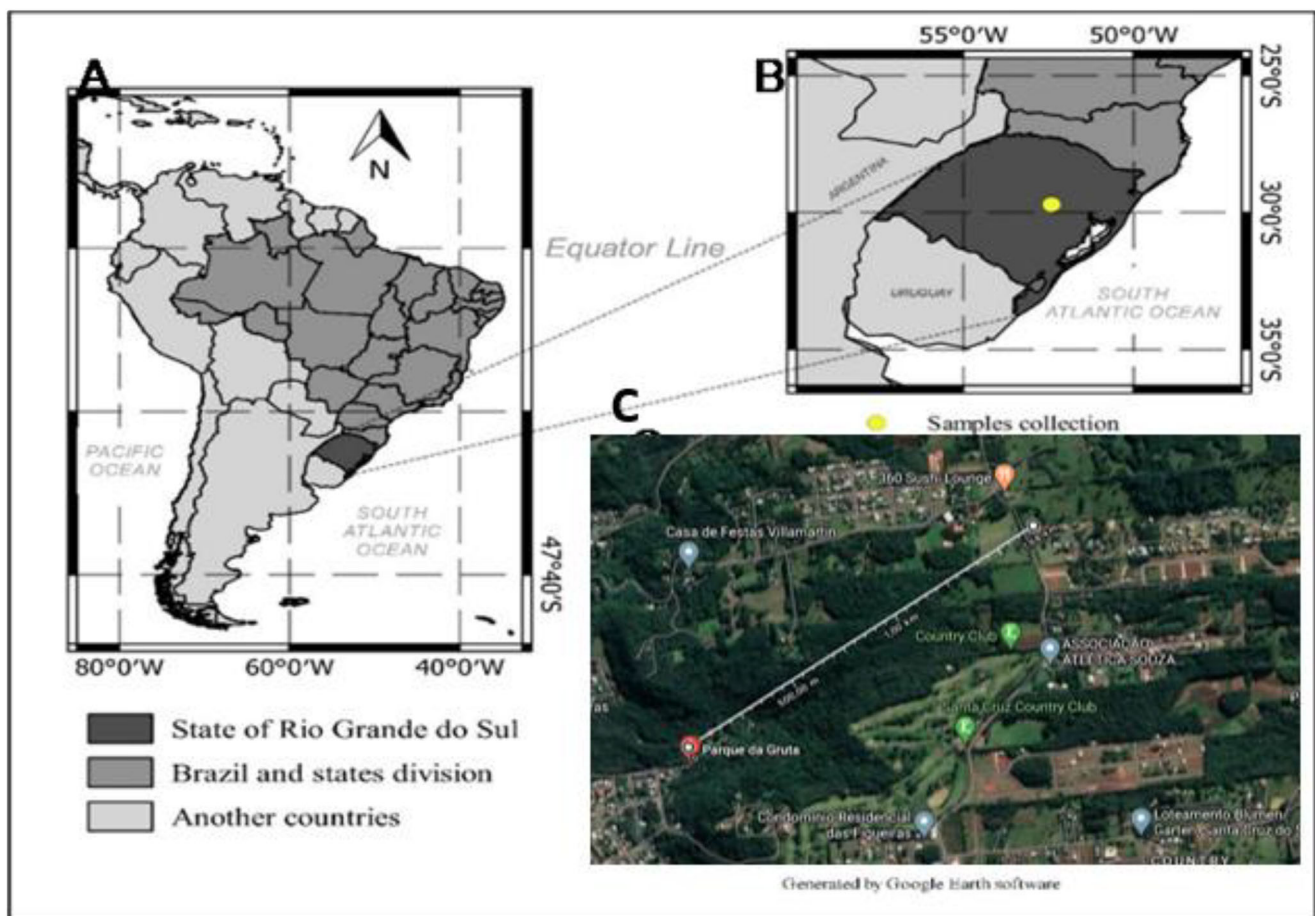
The purpose of this study is to access the viral sequences obtained from oral swabs of *Sapajus nigritus* (Rylands, 2012) [9]. This research will improve the restricted knowledge regarding virus diversity in non-human primates and will help to visualize which viruses are part of oral cavity and to understand their

connections and interfaces with the environment. Despite the lack of information and previous knowledge that wild animals are a source of potential zoonotic viruses, the data support a first viral description in oral swabs of capuchin monkeys. This study is relevant to describe and serve as a basis for further researches characterizing viruses in wildlife fauna.

## Material and methods

### Study area

The present study was conducted in 2017 in Santa Cruz do Sul (29° 43' 04" S 52° 25' 33" O), a municipality of Rio Grande do Sul state, Brazil (Fig. 1A). Figure 1B shows the point where samples were collected. Finally, Fig. 1C shows an image extracted from Google Earth, showing the Atlantic Forest fragment where capuchin monkeys were captured. This local is a park, situated 2 km in distance from the center of city. The area contains 17.4 ha, including caves, trails open to tourists on the native forest, and waterfalls. The area is inhabit for small rodents, bats, variety of birds, insects, fish,



**Fig. 1** Map of samples' collection. (A) Brazil in South America. (B) Rio Grande do Sul, South of Brazil. (C) Real image extracted from Google Earth showing the distance between the collection point and a residential point

crustaceans, small lizards, and snakes. Such animals have a frequent and close contact with humans who visited the local as tourist attraction.

### Sample collection

Free ranging, wild and adult capuchin monkeys (*S. nigritus*,  $n = 5$ ) were captured using Tomahawk-type cages, and the details of sampling is provided in Table 2 (Supplimentary Material). Following, the animals were submitted to anesthesia with ketamine (100 mg/mL) and xylazine (20 mg/mL) as previously described by Miranda et al. (2011). After 20 min of post-capture a veterinarian checked the animal welfare. These procedures were applied in order to avoid pain or distress during manipulation and sample collection. These procedures were applied in order to avoid pain or distress during manipulation and sample collection. The swabs were collected individually in oral cavities and stored in sterile tubes with 2 mL of Minimal Essential Medium (MEM) supplemented with 0.5 mg/mL penicillin-streptomycin (SIGMA). The tubes were kept on ice and stored at  $-80\text{ }^{\circ}\text{C}$ . Sedated animals were monitored closely during procedures and observed after release until regaining full control of movements. All animals successfully recovered from the anesthesia and were released in the same locality of capture after the procedure. All samples were collected in strict accordance with the Instituto Brasileiro do Meio Ambiente e Recursos Naturais (IBAMA) and Centro Epidemiológico de Vigilância em Saúde (CEVS), and under a license emitted by SISBIO/Ministry of the Environment, in accordance with the Brazilian authorization and conservation of biodiversity (License number: 56540-1).

### Viral concentration, nucleic acid extraction, and library preparation

Samples were gently vortexed for 2 min in their collection tubes. Following, 1 mL of each sample was clarified by centrifugation for 5 min at  $12000\times g$  in a microcentrifuge tube and then pooled into a single 15-mL conic tube. Next, the pool was filtered through a  $0.45\text{-}\mu\text{m}$  Millipore syringe filter® and the filtered suspension was submitted to ultracentrifugation at  $33000\times g$  for 3 h at  $4\text{ }^{\circ}\text{C}$  on a 25% sucrose cushion. The viral pellet was suspended in 500  $\mu\text{L}$  of ultrapure water, treated with nucleases (100 U of DNase I recombinant Roche® and PureLink 100 U of RNase A Thermo Fisher Scientific®) and submitted to nucleic acid extraction (UltraPure Buffer-Saturated Phenol, Life Technologies® or TRIzol® according to standard procedures). The quality of DNA extraction was checked in a 0.7% agarose gel, and the amounts of both nucleic acids obtained were measured using Qubit®. The extracted DNA was randomly amplified with phi29 DNA polymerase 10 U/ $\mu\text{L}$  (Thermo Fisher Scientific®) and the RNA was amplified using the REPLI-g

Cell WGA & WTA Kit QIAGEN®, both according the manufacturer's instructions. The final amplification products were checked in a 0.8% agarose gel and submitted to AMPure XP magnetic bead purification. After purification, the extractions were quantified and normalized to 2 ng/ $\mu\text{L}$  in a total of 20  $\mu\text{L}$ . The library was constructed using the Nextera XT DNA Library Prep Kit® ( $2\times 150$  bp), which simultaneously fragmented and tagged the adapters in a single enzymatic reaction. The sequences were generated by Illumina® MiSeq sequencer.

### Bioinformatics analysis

#### Assembly and Blast—first check

Reads were de novo assembled using MIRA [10] and SPAdes 3.10 assembler. The data generated were blasted using BLAST+ against the UniprotKB database and Viral database from NCBI and manually reviewed, using a cutoff e-value of 0.001. All the data are available on National Center of Biotechnology Information (NCBI): Bioproject SUB6191473 entitled “*Sapajus nigritus* raw sequence reads” coding a Sequence Read Archive (SRA) SUB6191751.

#### FrRIDA pipeline

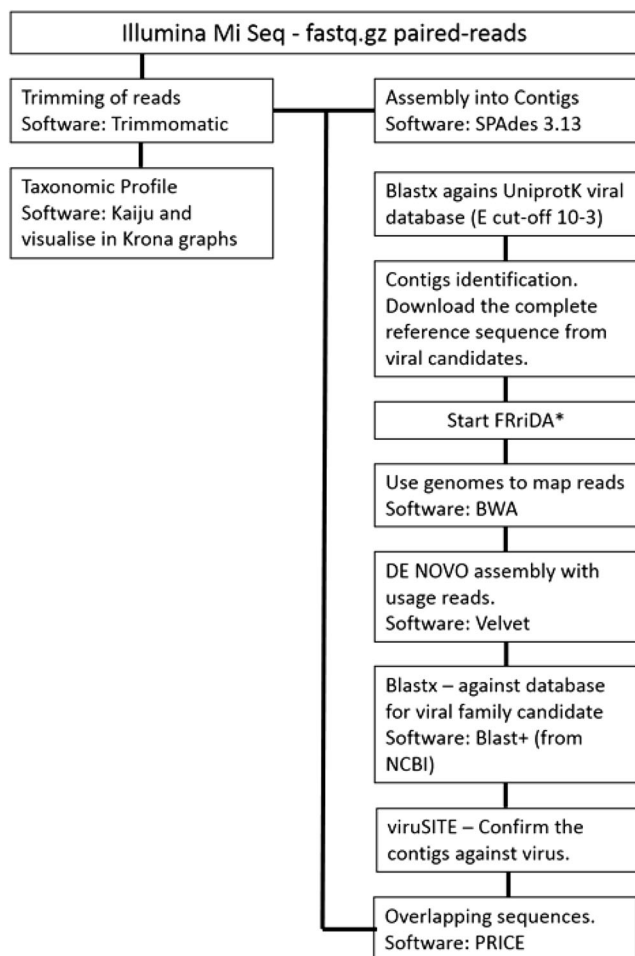
After checking the previous matches against viral sequences, each contig was analyzed individually. We downloaded the database from a viral family previously identified (ViPR) and mapped all the reads against a genome reference. After mapping the reads, the matched reads were used to produce new contigs (Velvet assembler). Then, we performed the BLASTx against the families and recovered the blasted contigs. We also used PRICE [11] for iterative contig extension, in order to obtain larger sequences. FrRIDA (Forced Raw reads in De novo Assembly) workflow, schematically represented in Fig. 2, was applied individually and used a Blast cutoff of 0.001 in order to create forced raw reads in de novo assembly and optimize the assembly of viral sequences.

#### Reads—taxonomy overview KAIJU

The interactive metagenomics visualization was applied using KRONA tools [12] to generate taxonomy profile graphs. We used the fastq.gz reads as input produce read-based estimation of taxonomical groups and further compare with our viral analysis.

#### Phylogenetic analysis

Sequence alignment of amino acid or nucleotide sequences of the conserved viral genes/proteins was carried using MAFFT [13]. Additional sequences were retrieved from GenBank after initial



**Fig. 2** FrRIDA pipeline, flowchart with the complete steps

Blast search. Maximum likelihood trees analyses were carried out using MEGA6®. Alignments were assigned to independent evolutionary models as shown in each subsequent phylogenetic tree.

### In silico validation—mapping the reads

In order to quantify and provide substantial data to confirm and support our findings, we checked the contigs and reads against reference genomes using a cutoff of 40% in the alignment (Supplemental material, Table 1). We performed the mapping using the BWA [14] tools and Geneious® package.

## Results

### Detection of viruses in monkey oral swabs—general data and taxonomic profiles

A total of 903,430,000 raw reads were generated from the oral swab samples in which 99% survived after trimming process (excluding the first nineteen nucleotides). After the quality checking, reads remained were assembled into 1612 contigs in length >

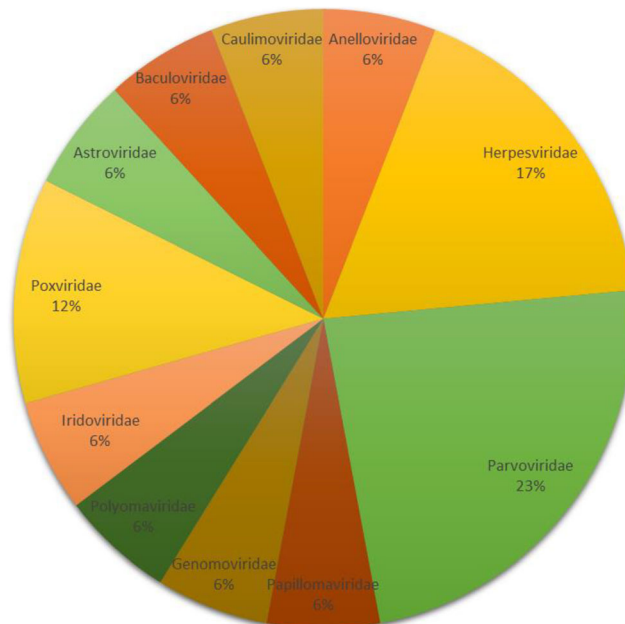
200 bp using SPAdes of total reads. The bioinformatics pipeline applied here detected 17 viral consensus sequences using 0.00915% of total reads, with significant identities to 11 families (Fig. 3). Viruses detected included single- and double-stranded DNA and RNA viruses, as well as enveloped and non-enveloped particles. Including complete genomes of *Genomoviridae*, *Anelloviridae*, and partial genes of *Baculoviridae*, *Astroviridae*, *Poxviridae*, *Iridoviridae*, *Polyomaviridae*, *Papillomaviridae*, *Parvoviridae*, *Caulimoviridae*, and *Herpesviridae* (Fig. 1 Supplementary Material).

Taxonomic profiling was carried out with Kaiju [15]. It generated the bubble graph displaying the number of classified reads against non-redundant databases as shown in Fig. 4. All the viral matches were checked using the e-value cutoff 0.001 as a parameter for significance (Table 1, Supplementary Material). Subsequently, we checked the reads to assemble the respective contigs (using bwa command), shown as coverage column on Table 1 in Supplementary Material. Among the assembled viral contigs, five were chosen to perform a more in-depth analysis. This included viral single- and double-stranded DNA sequences (linear and circular) that had a high coverage and quality fragment as previously described in methods.

### Viral diversity—overview

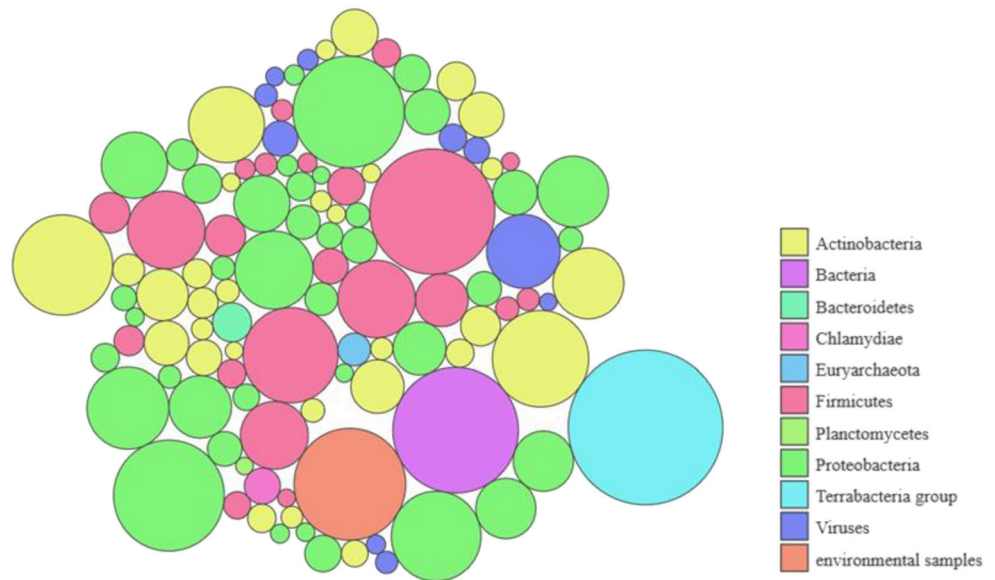
#### Anelloviruses

Anelloviruses are non-enveloped virions with a circular, single-stranded DNA genome of about 2.1 and 3.9 kb long. A complete viral genome, 3318 nucleotides long, was



**Fig. 3** Consensus sequences distribution of each viral family. The graphic was produced using Excel® office package. Data used is available in the Supplementary Material

**Fig. 4** Total of raw reads matched against the non-redundant protein database. The alignment provided 35% of total matched reads using the KAIJU Web Server



assembled using a total of 2639 reads in a maximum coverage of 281.1. According to the demarcation criteria used for anelloviruses, a range divergence of 56% and 35% of entire ORF 1 is necessary to identify a new genus or species, respectively. The sequence identified here displayed more than 40% divergence (56.4% of identity) with *Torque teno tamarin virus* (NC\_014085.1) which is consistent with a new species of anellovirus. The phylogenetic analyses using 63 ORF1 sequences (Fig. 5) show that the virus identified here, *Torque teno sapajus virus* (TTSapV), is placed in the same clade of *Simian torque teno virus*, among the *Epsilon torque viruses* (represented the contig 1 in Fig. 5).

### Genomovirus

Viruses belonging to the *Genomoviridae* family infect eukaryotes and contain a non-enveloped particle with a circular single-stranded DNA molecule of about 2.2 kb. This family includes nine genera, and its members have been identified from various environmental-, plant-, and animal-associated samples, demonstrating their abundance and ubiquity. So far, only one member has been cultivated, the only ssDNA virus known to infect a fungus (*Sclerotinia sclerotiorum*) [16]. Our analyses produced one initial contig, 246 nt long, which was submitted to a paired-read iterative contig extension using PRICE. Such extension generated a single molecule of 2239 nt, which corresponds to the complete genome of a new genomovirus, tentatively named *Sapajus gemycircularvirus* (Fig. 6A and B). Conservative motifs, the putative capsid protein, and Rep amino acid sequences were predicted by transferring annotations from GenBank database (Fig. 6A). The capsid sequences of genomoviruses are significantly more divergent than Rep sequences; thus, Rep is used for taxonomic analyses, as already described [17]. The

alignment of Rep A showed 69% of identity with *Genomoviridae* sp. (QCW23704.1), isolated from mouse tissue (unpublished data). Our finding is in accordance with the species demarcation cutoff, which preconizes a minimum of 78% genome-wide pairwise identity. A particularity of this genome is an intergenic region, which is potentially a loop structure with a nanonucleotide motif, an important skill for viral rolling-circle replication (Fig. 6A).

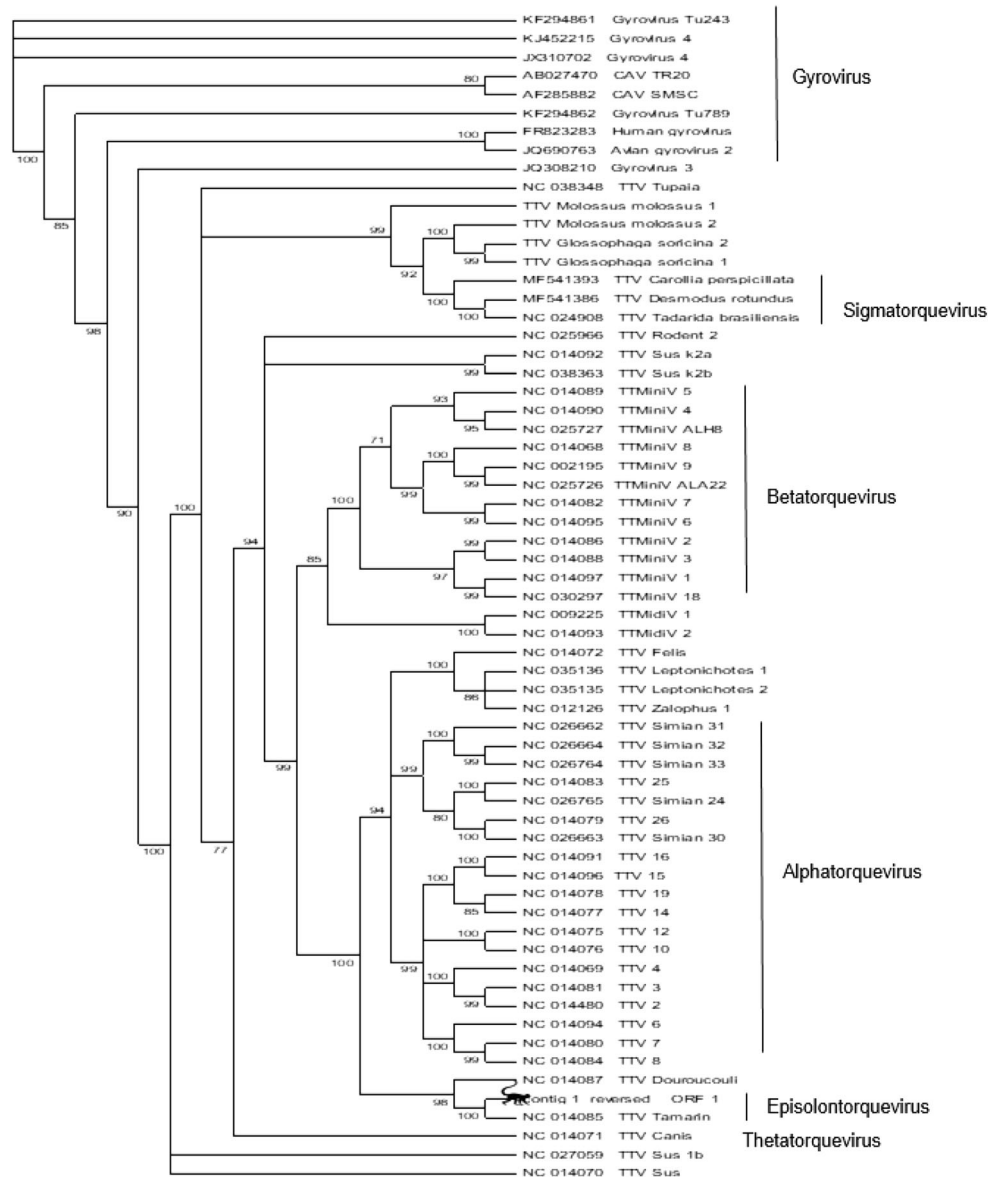
### Herpesviruses

Members of the *Herpesviridae* are enveloped virions with a double-stranded DNA genome of about 160–200 kbp. The subfamily *Gammaherpesvirinae* comprises 4 genera: *Macavirus*, *Percavirus*, *Rhadinovirus*, and *Lymphocryptovirus*, and some isolates were already identified in samples from non-human primates [18]. We assembled 3 contigs which fall in the genus *Gammaherpesvirus*, as shown in Table 1, Supplementary Material. From these, two display more than 75% identity with *Human gammaherpesvirus 4* and one displays 52% identity with *Callitrichine gammaherpesvirus*. A phylogenetic tree was built fitted by JTT + G model to reinforce the findings among primate gammaherpesviruses (Fig. 7) using a partial sequence of the ssDNA binding region from contig ID 861 (named *Sapajus* partial in the subsequent phylogenetic tree below).

### Polyomavirus

*Polyomaviridae* is a virus family that includes non-enveloped viruses of about 4.5–5.0 or 5.5 Kbp long of circular, double-stranded DNA. Here, we obtained one consensus sequence, covering part (876 bp) of the VP1 gene. Such contig displays 78% of identity with *Otomops polyomavirus*, with a

**Fig. 5** Maximum likelihood tree of the complete ORF1 of anelloviruses. Nucleotide sequence of ORF 1 was compared with 62 anelloviruses from mammals and birds, retrieved from GenBank. The substitution model estimated was GTR + G with 1000 replicates

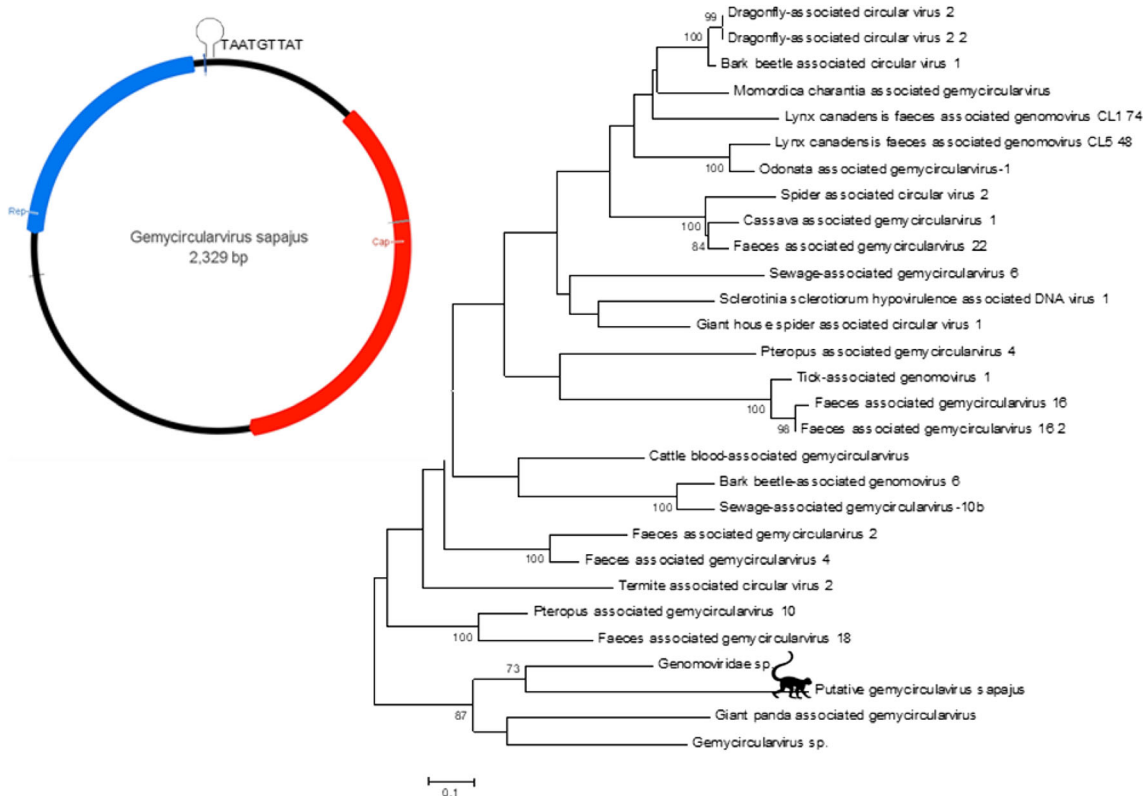


maximum coverage of 8 (Q20 of 88.5%). The phylogenetic analyses were supported by the substitution model HKY85 with a calculated gamma variable (0.864). The phylogenetic analyses place the newly identified polyomavirus in the same clade of two other non-human primate (chimpanzee and African gray monkey)-related polyomaviruses (Fig. 8).

### Parvoviruses

Members of the *Parvoviridae* are non-enveloped viruses containing a linear single-stranded DNA genome. We obtained four assembled contigs (from 12 and 20 in coverage), with significant identity to members of this family. One contig shows identity to members of the subfamily *Parvoviridae* (which includes vertebrate-infecting viruses) and the other

three display identity of 37–55% with members of the subfamily *Densovirinae*, which infects arthropods. The 1152 nt contig comprises 72% of the complete gene from ORF3, showing 52% identity to *Parus major densovirus*, a densovirus isolated from birds [19]. The 789 nt contig consists of 33% of the complete VP1 and exhibits 41% identity with the viral polypeptide from a parvovirus previously described in *Lepidoptera*. In addition, 51% of structural protein from *Ambidensovirus* was detected, displaying 37% of identity to this genus. The partial NS1, conserved between parvoviruses and preconized for analysis, was used to build a phylogenetic tree recovered from contig ID 597 (Table 1, Supplementary Material). The phylogenetic tree (Fig. 9) indicates that the assembled sequence (named putative NS1 Sapajus) is related to members of the *Densovirinae* subfamily.



**Fig. 6** (A) Schematic representation of the genome of the *Sapajus gemycircularvirus*. In blue: Rep. In red: Cap. The sequence TAATGTTAT is a conservative motif. (B) Maximum likelihood tree of

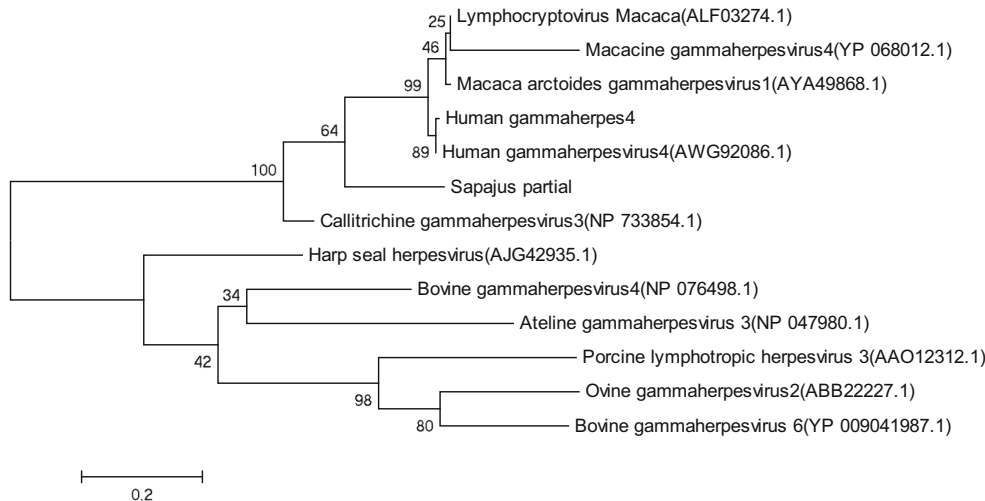
29 Rep catalytic A domain of unclassified gemycircularviruses. The substitution model estimated was LG + G + I with 1000 replicates

**Additional viruses**

In addition to the sequences mentioned above, other virus-related sequences were retrieved. Such reads did not allow further phylogenetic analyses, as they are short and did not correspond to species or genus demarcation genes. Among these, we found a 420 pb consensus sequence with 95% identity to a hypothetical protein from *human*

*papillomavirus* 4 and two consensus sequences which display significant identity to poxviruses; one contig displayed identity of to an insect-infecting poxvirus and the other has identity of 52% with *Tatera poxvirus*, which has been associated to rodents in Africa [20] (Table 1, Supplementary Material). In addition, contigs related to *Iridoviridae*, *Astroviridae*, *Caulimoviridae* and *Baculoviridae* were retrieved (Table 1, Supplementary Material).

**Fig. 7** Partial analysis of the herpesvirus ssDNA binding amino acid sequences of five gammaherpesviruses from human and non-human primates. The phylogenetic analysis was supported by JTT + G model with 1000 replicates







## Discussion

The analysis of the oral virome of capuchin monkeys by metagenomics allowed the detection of both DNA and RNA viral genomes, classified in eleven families. The majority of the virus-associated sequences (72% of contigs) belong to bacteriophages, meaning *Siphoviridae*, *Myoviridae*, and *Podoviridae* families (data not shown). The assembled sequences analyzed here corresponded to vertebrate (host and non-host-related), invertebrate, and plant viruses. Despite the lack of a common sequence among viruses, we used laboratory procedures to minimize non-virus nucleic acids (as described in methods) and preserve as much as possible viral nucleic acids and analyzed the data using our *in-house* pipeline (FrRIDA). We had difficulties to amplify high numbers of viral reads, as shown in a bubble graph representing a reads-based result. Previously trimmed reads were examined in distinct assemblers (data not shown) and matched against databases resources (viPR, viruSITE, Expaty, NCBI, NCBI-virus) using nucleotides and translated nucleotide sequences. The application of FrRIDA pipeline utilizing viruSITE as database [21] provided fitted consensus sequences attending the quality parameters.

Despite the efforts to enrich the viral sequences from the oral swabs, only 0.2% of the obtained sequences were used to assemble viral contigs. We analyzed oral swabs from a group of juvenile, apparently healthy, wild *S. nigritus* ( $n = 5$ ) (see support data about each sample in the Supplementary Material). After trimming and assembly, seventeen consensus sequences were identified, which were related and showed variable identities to eleven viral families including *Anelloviridae*, *Astroviridae*, *Parvoviridae*, *Herpesviridae*, *Genomoviridae*, *Papillomaviridae*, *Poxviridae*, and *Polyomaviridae*.

Viruses belonging to the *Anelloviridae* family, also ssDNA/circular, are widely distributed worldwide. Nevertheless, neither the type of interaction with their hosts nor their direct association with the etiology of specific diseases is entirely understood [22]. Previously, studies suggested the association with pathological conditions, such as hepatitis, cancer, and respiratory diseases as well as hematological and autoimmune disorders [23]. A study conducted with 17 African captive primate species reported 100% of prevalence of anelloviruses [24]. Our study allowed the assembly of a complete anellovirus genome tentatively named as *Torque teno sapajus virus* (TTSapV). Such sequence is, according to the demarcation criteria of anelloviruses, classified in the genus *Epsilon Torque teno virus*, and clustered with a *Tamarin torque teno virus* identified in 2000 by Abe and collaborators [25]. Both viruses are significantly distinct from human anellovirus isolates (see Fig. 5).

Circular single-stranded DNA viruses (ssDNA/circular) infect all the three domains of life and are associated to the development of clinical disease, environmental, and economic

losses [26]. *Gemycircularvirus* is the most abundant genus among the *Genomoviridae* family. Previous metagenomics studies revealed their Ubiquiti distribution, from mammals, including primates [16], to unicellular algae [27]. We proposed a new species among the family named *Sapajus*-associated *gemycircularvirus*, be aware of its extent presence.

The present study revealed the presence of sequences which share identity to *human gammaherpesvirus 4* and *Callitrichine herpesvirus*, both are classified in the *Gammaherpesvirinae* subfamily. Gammaherpesvirus infections are frequent in different species, and occasionally, they may be related to the occurrence of cancer [28]. However, the role of these viruses in diseases of non-human primates is not clear. Primates may be naturally infected by several herpesvirus, which emerged about roughly 400 million years ago [29]. *Macacine alphaherpesvirus* is, for instance, related to asymptomatic infections in its natural host (macaques); however, when transmitted to non-native hosts, it may cause a fatal disease [30]. Here, all animals submitted to oral swab collection were apparently healthy, suggesting they were developing an asymptomatic infection.

Here, we identified a partial sequence of a VP1 gene which shares 78% identity to *Otomops polyomavirus* (YP\_007346983.1). This virus was described originally from the free-tailed bat from the African continent. The first polyomavirus found in new world primates was described in 2008, when Ernts and collaborators identified an isolate from *Saimiri* sp., putatively named *squirrel monkey polyomavirus*. Later, polyomaviruses from both Old and New World monkeys were further described [31]. Mammalian polyomaviruses are highly host-specific and interspecies transmission is considered sporadic or non-existent. The sequence described here may be either related to food sources or may be host-specific; however, further studies should be performed to answer why such sequence was found among these animals.

Capuchin monkeys are considered opportunistic omnivores, and they feed on a wide range of food types, including fishes and crustaceans. Among the viruses found from these samples are crustacean/fish-related viruses from families *Parvoviridae* and *Iridoviridae*, which suggests a contact of these animals with aquatic animals. In addition, we detected insect-related parvovirus and baculovirus, and a bird-related parvovirus. Such sequences probably originate from feeding, corroborating with the wide range of food sources observed elsewhere [32]. As an important part of their diet, fruits are consumed from natural resources. We found a *Caulimoviridae*-related sequence; such viruses cause several diseases in productive grapevines.

The astrovirus and polyomavirus sequences amplified here display significant identities to bat-related viruses. A partial sequence of the non-structural polyprotein displays 42% identity to the bat bastrovirus-like (YP\_009333174.1), a sequence recovered from feces of bats collected in Vietnam

(unpublished data). Astroviruses are generally associated with diarrhea; however, they may be associated to nephritis, hepatitis, and encephalitis or even asymptomatic infections. Bats are recognized to harbor a variety of astroviruses which infect exclusively bats and have not been found in any other mammal [33]. Such animals may also be part of the diet of non-human primates, as non-usual, probably leading to the interpretation that capuchin monkeys had a contact with bat feces that could be contaminated by the virus. These animals share the same water resources and trees to live and to explore the fruits, leading a near contact with probably biological fluids of each other. These interactions can explain the detection of such sequence as part of the oral virome described here.

We identified a contig that displays 95% identity to human papillomavirus 4, which might indicate an association with virus infections in these animals, similar to what has been described from howler capuchin monkeys (*Alouatta fusca*) [34]. However, lesions in the oral mucosae were not observed, and these viruses may thus be part of normal microbial content, as described previously for other species [35, 36].

Metagenomic studies applied to the description of virome of different species are expanding our knowledge on the diversity and distribution of viruses in nature. Here, about 50% of the generated data did not display significant identity to any virus-related sequence available on databases; however, we unveil important information on the structure of the *Sapajus nigritus* oral virome. Further studies should be performed in order to improve and deepen our knowledge on *Sapajus* sp. virome and microbiome, allowing us to better understand the virus circulation among these animals and their role in the development of host-specific diseases.

## Conclusions

This study contributes to access the viral sequences from capuchin monkeys, leading to improve our restrict knowledge regarding microbial communities in non-human primates.

**Acknowledgments** We thank Dr. Fabiana Quos Mayer from Instituto de Pesquisas Veterinarias Desiderio Finamor (IPVDF) for providing the sequencing service. We also thank Ms. Rai Nunes dos Santos who built the map on QGIZ®. The correspondence author would like to thank the Animal and Plant Health Agency (United Kingdom government) for providing the access to Scientific Environmental Computer and bioinformatics group collaboration. We emphasize the scientific collaboration and would like to thank especially Dr. Denise A Marston, Dr. Daniel Dorey-Robson, and Dr. Lorraine McElhinney, head of the APHA Wildlife Zoonoses & Vector-Borne Diseases Research Group.

**Funding information** The correspondence author would like to thank CAPES process number (PDSE 88881.198088/2018-01) for financial support.

## References

- Dunn JC, Cristóbal-Azkarate J (2016) New World monkeys. Nature Education Knowledge 7(6):1
- IUCN, Kierulff MCM, Mendes SL, Rylands AB (2015) *Sapajus nigritus*. In: The IUCN Red List of Threatened Species 2015, p e.T136717A70614145
- Wolfe ND, Escalante AA, Karesh WB, Kilbourn A, Spielman A, Lal AA. Wild primate populations in emerging infectious disease research: the missing link? [Internet]. Vol. 4, Emerging Infectious Diseases. 1998. p. 149–158
- Dewhurst FE, Klein EA, Thompson EC, Blanton JM, Chen T, Milella L et al (2012 [cited 2018 Nov 13]) The canine oral microbiome. PLoS One 7(4):e36067
- Pride DT, Salzman J, Haynes M, Rohwer F, Davis-Long C, White RA et al (2012 May [cited 2019 Aug 23]) Evidence of a robust resident bacteriophage population revealed through analysis of the human salivary virome. ISME J 6(5):915–926
- Willner D, Furlan M, Schmieder R, Grasis JA, Pride DT, Relman DA, Angly FE, McDole T, Mariella RP, Rohwer F, Haynes M (2011) Metagenomic detection of phage-encoded platelet-binding factors in the human oral cavity. Proc Natl Acad Sci U S A 108(SUPPL. 1):4547–4553
- Bodewes R, Ruiz-Gonzalez A, Schapendonk CM, Van Den Brand JM, Osterhaus ADME, Smits SL (2014 [cited 2018 Aug 9]) Viral metagenomic analysis of feces of wild small carnivores. Virol J [Internet] 89(1):11
- Nelson KE (2015 [cited 2018 Oct 27]) An update on the status of current research on the mammalian microbiome. ILAR J 56(2): 163–168
- Rylands AB, Mittermeier RA, Silva JS Jr (2012) Neotropical primates: taxonomy and recently described species and subspecies. Int Zoo Yearb 46:11–24. <https://doi.org/10.1111/j.1748-1090.2011.00152>
- Chevreur B, Pfisterer T, Drescher B, Driesel AJ, Müller WEG, Wetter T, Suhai S (2004) Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Res 14(6):1147–1159
- Ruby JG, Bellare P, DeRisi JL (2013) PRICE: software for the targeted assembly of components of (Meta) genomic sequence data. G3&#58; Genes|Genomes|Genetics 3(5):865–880
- Ondov BD, Bergman NH, Phillippy AM (2011) Interactive metagenomic visualization in a Web browser. BMC Bioinformatics 12:385
- Katoh K, Standley DM (2016) A simple method to control over-alignment in the MAFFT multiple sequence alignment program. Bioinformatics. 32(13):1933–1942
- Li H, Durbin R (2009) Making the leap: Maq to BWA [Internet]. In: Mass Genomics, vol 25
- Menzel P, Ng KL, Krogh A (2016 Apr) Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun 13:7
- Krupovic M, Said •, Ghabrial A, Daohong Jiang •, Varsani A. Genomoviridae: a new family of widespread single-stranded DNA viruses. Arch Virol 2016;161
- Varsani A, Krupovic M (2017) Sequence-based taxonomic framework for the classification of uncultured single-stranded DNA viruses of the family Genomoviridae. Virus Evol 3:1
- Herpesviridae - dsDNA Viruses - dsDNA Viruses (2011) - International Committee on Taxonomy of Viruses (ICTV). [cited 2019 Aug 23]. Available from: [https://talk.ictvonline.org/ictv-reports/ictv\\_9th\\_report/dsDNA-viruses-2011/w/dsDNA\\_viruses/91/herpesviridae](https://talk.ictvonline.org/ictv-reports/ictv_9th_report/dsDNA-viruses-2011/w/dsDNA_viruses/91/herpesviridae)
- Yang W-T, Shi S-H, Jiang Y-L, Zhao L, Chen H-L, Huang K-Y et al (2016 Jul) Genetic characterization of a densovirus isolated

- from great tit (*Parus major*) in China. *Infect Genet Evol* [Internet] 41:107–112
20. Lourie B, Nakano JH, Kemp GE, Setzer HW (1975) Isolation of poxvirus from an African Rodent [Internet]. *J Infect Dis. Oxford University Press* 132:677–681
  21. Stano M, Beke G, Klucar L. *viruSITE-integrated database for viral genomics. Database (Oxford)*.2016
  22. Brajão de Oliveira K (2015) Torque teno virus: a ubiquitous virus. In: *Revista Brasileira de Hematologia e Hemoterapia*, vol 37. Elsevier Editora Ltda, pp 357–358
  23. Spandole S, Cimponeriu D, Berca LM, Mihăescu G Human anelloviruses: an update of molecular, epidemiological and clinical aspects. *Arch Virol* 160(4):893–908
  24. Leary TP, Erker JC, Chalmers ML, Desai SM, Mushahwar IK (1999) Improved detection systems for TT virus reveal high prevalence in humans, non-human primates and farm animals. *J Gen Virol* 80(8):2115–2120
  25. Abe K, Inami T, Ishikawa K, Nakamura S, Goto S (2000) TT virus infection in nonhuman primates and characterization of the viral genome: identification of simian TT virus isolates. *J Virol* 74(3): 1549–1553
  26. Phan TG, Kapusinszky B, Wang C, Rose RK, Lipton HL, Delwart EL. The fecal viral flora of wild rodents. *PLoS Pathog.* 2011 Sep;7(9)
  27. Yoon HS, Price DC, Stepanauskas R, Rajah VD, Sieracki ME, Wilson WH, Yang EC, Duffy S, Bhattacharya D (2011) Single-cell genomics reveals organismal interactions in uncultivated marine protists. *Science*. 332(6030):714–717
  28. Estep RD, Wong SW (2012) Nonhuman primate gamma-herpesviruses and their role in cancer. *Curr Cancer Res* 8:201–214
  29. Grose C (2012) Pangaea and the out-of-Africa model of varicella-zoster virus evolution and phylogeography. *J Virol* 86(18):9558–9565
  30. Eberle R, Jones-Engel L. *J EMERG DIS VIROL. Understanding Primate Herpesviruses*. 2017
  31. Scuda N, Madinda NF, Akoua-Koffi C, Adjogoua EV, Wevers D, Hofmann J, et al. Novel polyomaviruses of nonhuman primates: genetic and serological predictors for the existence of multiple unknown polyomaviruses within the human population. *PLoS Pathog.* 2013 Jun;9(6)
  32. Hohmann G. The Diets of non-human primates: frugivory, food processing, and food sharing. In: Hublin J-J, Richards MP, editors. *The evolution of hominin diets: integrating approaches to the study of palaeolithic subsistence* Dordrecht: Springer Netherlands; 2009. p. 1–14
  33. Chu DKW, Poon LLM, Guan Y, Peiris JSM (2008 Sep) Novel astroviruses in insectivorous bats. *J Virol* 82(18):9107–9114
  34. Silvestre RVD, Souza AJS de, Júnior ECS, Silva AK, Mello WA de, Nunes MRT, et al. First New World primate papillomavirus identification in the Atlantic Forest, Brazil: *Alouatta guariba papillomavirus 1*. *Genome Announc.* 2016. 4(4)
  35. Antonsson A, Forslund O, Ekberg H, Sterner G, Hansson BG (2000 Dec) The ubiquity and impressive genomic diversity of human skin papillomaviruses suggest a commensalic nature of these viruses. *J Virol* 74(24):11636–11641
  36. Rector A, Ranst M Van. *Animal papillomaviruses*. *Virology*. 2013. 445(1–2)

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