

Metagenomic analysis reveals novel dietary-related viruses in the gut virome of marmosets hybrids (*Callithrix jacchus* x *Callithrix penicillata*), Brazil

Thamiris dos Santos Miranda ^{a,1}, Francine Bittencourt Schiffler ^{a,1}, Mirela D'arc ^a, Filipe Romero Rebello Moreira ^{a,b}, Matheus Augusto Calvano Cosentino ^a, Amanda Coimbra ^a, Ricardo Mouta ^a, Gabriel Medeiros ^a, Déa Luiza Girardi ^a, Victor Wanderkoke ^a, Caique Ferreira Amaral Soares ^{c,d}, Talitha Mayumi Francisco ^{c,d}, Malinda Dawn Henry ^d, Bianca Cardozo Afonso ^{c,d}, Flávio Landim Soffiati ^c, Suelen Sanches Ferreira ^c, Carlos Ramon Ruiz-Miranda ^{c,d}, Marcelo Alves Soares ^{a,e}, André Felipe Andrade dos Santos ^{a*}

^a Laboratório de Diversidade e Doenças Virais, Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, RJ, Brazil

^b Department of Infectious Disease Epidemiology, Imperial College London, London, UK.

^c Associação Mico-Leão-Dourado, Silva Jardim, Rio de Janeiro, RJ, Brazil

^d Laboratório de Ciências Ambientais, Universidade Estadual do Norte Fluminense Darcy Ribeiro (UENF), Campos dos Goytacazes, RJ, Brazil

^e Programa de Oncovirologia, Instituto Nacional de Câncer, Rio de Janeiro, RJ, Brazil

* Corresponding author:

e-mail: afsantos.ufrj@gmail.com (AFAS; <https://orcid.org/0000-0003-3372-1552>)

¹ These authors contributed equally to this work

Abstract

Viral metagenomics has contributed enormously to the characterization of a wide range of viruses infecting animals of all phyla in the last decades. Among Neotropical primates, especially those free-living introduced, knowledge about viral diversity remains poorly studied. Therefore, through the use of metagenomics based on virus enrichment, we explored the viral microbiota present in the feces of introduced common marmosets (*Callithrix sp.*) in three locations from the Silva Jardim region in the State of Rio de Janeiro, Brazil. Fecal samples were collected from nine marmosets, pooled into three sample pools and sequenced on Illumina MiSeq platform. Sequence reads were analyzed using a viral metagenomic analysis pipeline and two novel insect viruses belonging to the *Parvoviridae* and *Baculoviridae* families were identified. The complete genome of a densovirus (*Parvoviridae* family) of 5,309 nucleotides (nt) was obtained. The NS1 and VP1 proteins share lower than 32% sequence identity with the corresponding proteins of known

members of the subfamily *Densovirinae*. Phylogenetic analysis suggests that this virus represents a new genus, named *Tritonambidensovirus* due to telomeric structures at the 3' and 5' ends of the genome. The novel species received the name *Fecalis tritonambidensovirus 1*. The complete circular genome of a baculovirus of 107,191 nt was also obtained, showing 60.8% sequence identity with the most closely related member of the *Baculoviridae* family. Phylogenetic analysis suggests that this virus represents a new species of *Betabaculovirus*, named *Callithrix fecalis granulovirus*. In addition, sequences from several families of arthropods in the three pools evaluated were characterized (contigs ranging from 244 to 6,750 nt), corroborating the presence of possible insect hosts with which these new viruses may be associated. Our study expands the knowledge about two viral families known to infect insects, an important component of the marmosets' diet. This identification in hosts' feces samples demonstrates one of the many uses of this type of data and could serve as a basis for future research characterizing viruses in wildlife using noninvasive samples.

Keywords: Free-living host, Feces, HTS, Invertebrate virus, Diet

1. Introduction

Recent advances in high throughput sequencing (HTS) technologies have allowed metagenomic analysis to reveal an increasing genetic diversity of viral genomes ranging from unicellular bacteria, archaea and protozoa to multicellular organisms such as plants, fungi, invertebrates and vertebrates (Temmam *et al.*, 2014). Millions of wild fauna viruses are poorly characterized or totally unknown to man (Carroll *et al.*, 2018). Among vertebrates, little is known about the virome of Neotropical Primates (NP), especially among free-living populations. ADuarte *et al.*, (2019) characterized the fecal virome of free-living *Sapajus libidinosus* in the Brazilian Cerrado and identified novel known and putative viruses. However, it is currently estimated that NP are distributed in 183 species, 21 genera and three families (*Atelidae*, *Pitheciidae* and *Cebidae*), according to molecular data (Dumas and Mazzoleni, 2017; Perelman *et al.*, 2011; IUCN/SSC Primate Specialist Group, 2021), for which there are no systematic studies of viromes. Characterizing viruses infecting this group, as well as their susceptibility to infectious diseases, is key for understanding primates' disease ecology, evolution and conservation.

The two species of common marmosets, the white-tufted (*Callithrix jacchus*) and the black-tufted (*Callithrix penicillata*) are diurnal, arboreal NP, native of northeastern and central Brazil, respectively (Malukiewicz *et al.*, 2021). Adults can weigh from 300 to 500 g and have an average life expectancy of 5 to 12 years in research colonies (Ross *et al.*, 2012). Marmosets live in family groups of up to 15 individuals with a dominant breeding pair (Digby, 1995). In the wild, their diet is mainly composed of plants, fruits, seeds, flowers, insects, and small animals (Pinheiro and Mendes Pontes, 2015). As a result of the illegal wildlife trade, these marmosets have been introduced to several regions outside and within Brazil, where they are considered a threat to endemic species (Pinheiro and Mendes Pontes, 2015). One of these regions is the Atlantic coastal forest of Rio de Janeiro state, Southeast Brazil. In this region, these marmosets occur as hybrids of both species, have been ecologically successful and are a threat to the entire population of endangered golden lion tamarins (*Leontopithecus rosalia*) (Malukiewicz *et al.*, 2015; Ruiz-Miranda *et al.*, 2006). The biology of marmosets is similar to lion tamarins, therefore they can readily compete for food, shelter and can transmit etiological agents not currently present in lion tamarins, becoming a concern for the conservation of the native species (dos Santos Sales *et al.*, 2010).

Analyses of the viral microbiota can be useful in efforts to preserve host biodiversity and to guide conservation strategies (Stumpf *et al.*, 2016). Feces might contain genetic material of pathogens, food sources (animals and/or plants), and commensal gut microbes, being an

important resource for the analysis of ecological networks and the discovery of new viruses (Ge *et al.*, 2012; van den Brand *et al.*, 2012). In addition, feces are a non-invasive sample material, providing a safe alternative for biological sampling (Bergner *et al.*, 2019), since the small body size of many NP is a limiting factor for the collection of other types of samples, such as blood. Moreover, feces samples collection causes little to no animal stress, an ethically desirable feature.

Therefore, in this study, we characterized for the first time the viral metagenome of free-living introduced hybrids of *C. jacchus* x *C. penicillata* using stool samples. A potentially new genus and consequently a new species of Densovirus (family *Parvoviridae*) and a new species of *Betabaculovirus* (family *Baculoviridae*) were discovered through the analysis of HTS data, provisionally named *Fecalis tritonambidensovirus* 1 (FtDV) and *Callithrix fecalis granulovirus* (CafeGV), respectively. Evolutionary analyses indicate the new viruses are closely related to viruses previously characterized in insects, suggesting the reported experiment captured viral diversity associated with marmoset arthropod diet.

2. Material and Methods

2.1. Sample Collection

Fecal samples were collected from nine common marmosets living in three localities of Silva Jardim municipality, in the state of Rio de Janeiro, Brazil (**Table 1**). The marmosets were accustomed to regular human contact and consistently monitored by the *Associação Mico Leão Dourado* (AMLD). They were captured individually with Tomahawk® live traps baited with bananas and placed on bamboo platforms positioned inside the forest fragment at 1.5 m above the ground. After capture, animals were transported to the laboratory of the AMLD, also in *Silva Jardim*, for sanitary evaluation and sample collection. In the laboratory, opportunistic fecal samples were collected on clean paper surfaces positioned under traps or directly collected from animals during manipulation. Animals were immobilized with injection of ketamine (approximately 10-15 mg/Kg of body weight) into the intramuscular region of the inner thigh and general information were collected for all sampled animals, such as: ID number, group, species, age, sex, weight, and clinical conditions. Each fecal sample was transferred to 15 mL falcon tubes with RNAlater™ (Thermo Fisher Scientific, Walham, USA), a nucleic acid preservation buffer, with a volumetric ratio of feces-to-buffer of roughly 1:1. Samples were kept at room temperature and shipped to the *Laboratório de Diversidade e Doenças Virais* (LDDV) at *Universidade Federal do Rio de Janeiro* (UFRJ), Rio de Janeiro, Brazil, to be stored at -20 °C. All procedures were performed by veterinarians and biologists from AMLD and all samples were collected following the national guidelines and provisions of CONCEA (National Council for

Animal Experimentation Control, Brazil), which included animal welfare standard operating procedures. This project was approved by the Ethics Committee on the Use of Animals (CEUA) of UFRJ (reference number 037/14). This study is reported in accordance to ARRIVE guidelines (<https://arriveguidelines.org/resources/questionnaire>).

2.2 High-Throughput Sequencing

To maximize sequencing efficiency, samples were pooled into three groups according to locality (Afetiva, Igarapé and Iguape) with four, two and three samples each, respectively (**Table 1**). The samples were vigorously vortexed until complete homogenization. Afterwards, approximately 1 mL of sample was transferred to the extraction bead tube (MP Biomedicals, Santa Ana, CA, USA) to disrupt the cells. The supernatant was collected by centrifugation at $6,000 \times g$ for 10 min at 4°C, followed by filtering using Millex-HV 0.45 µm SLHV0135L (Merck & Co., Radington Township, NJ, USA) to remove host cells. The filtrates were submitted to an ultracentrifugation at 35,000 g for 90 minutes at 4°C to separate the enriched viral background fraction. The supernatant was discarded and approximately 0.2 mL of the sample was incubated with DNases (Promega and Epicentre, Madison, WI, USA), RNase A (Ambion, USA) and Benzonase (Sigma, USA) for 90 minutes at 37°C to digest unprotected nucleic acid. Viral nucleic acids were then isolated using QIAamp MinElute Virus Spin Kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions with some modifications described (Cosentino *et al.*, 2022). Complementary DNA (cDNA) synthesis was performed using the Superscript III First Strand Synthesis Supermix Kit (Thermo Fisher Scientific) and for the synthesis of the second strand of cDNA was used the Klenow fragment 3'-5' exo (New England Biolabs Inc., Ipswich, MA, USA), following the manufacturer's instructions. Mixed nucleic acid (DNA and cDNA) from all samples were subjected to library construction using Nextera XT DNA Sample Preparation Kit (Illumina Inc., San Diego, USA) and sequenced on an Illumina MiSeq platform (Illumina) at the Department of Genetics, UFRJ.

2.3 Bioinformatics Analyses

Sequencing data was processed with a custom pipeline. Raw reads were filtered with Fastp v.0.20.1 (Chen *et al.*, 2018), which removed adapters and low-quality reads (Phred score < 30 and fragment size < 50 bp). Contigs were *de novo* assembled with the Meta-spades software v.3.15.3 (Nurk *et al.*, 2017). Both reads and contigs from each sample were taxonomically assigned using Diamond v.2.0.14 (Buchfink *et al.*, 2015), with an e-value cut-off of 10^{-5} and parameters '--more-sensitive' and '--max-target 1' specified. The similarity search was

performed against the complete non-redundant protein database from NCBI (as of July 27th, 2021). Interactive visualization plots were generated with Krona v.2.7.1 (Ondov *et al.*, 2011). Following (Farouni *et al.*, 2020), to avoid false positive results due to sequencing index-hopping (incorrect assignment of reads to a given sample), we considered as invalid the identification of a viral family when it was detected by a number of reads that was less than 1% of the highest count identified for the same family among all other samples. Sequences from samples that fell below this threshold were not considered for family-specific downstream analysis.

2.4 Genome assembly, annotation and secondary genome structure prediction

Preliminary taxonomic assignment results indicated the presence of two novel near complete genomes from distinct viral families (*Parvoviridae* and *Baculoviridae*). To investigate whether complete genome assemblies could be obtained from the generated data, an extensive set of analysis was performed with the Geneious software v2021.2.2 (Kearse *et al.*, 2012).

For the *Parvoviridae* family, raw reads were mapped to the contig obtained through *de novo* assembly, in order to complete the genome of the virus. The Geneious mapper was used with the highest sensitivity setting. The limits of the new consensus genome sequence were determined through the prediction of the secondary structure of the Inverted Terminal Repeats (ITRs), characteristic of this viral family, using the Vienna package RNAfold tool v. 2.5.1. (Lorenz *et al.* 2011), built in the Geneious software. Both ITR sequences were also aligned using the Muscle software v. 3.8.425 (Edgar, 2004) and had their sequences and secondary structures visually compared to determine their status as homo or hetero-telomeric. Through the EMBOSS Nucleotide Analysis v. 1.1.1 Geneious plugin, the transcription factors associated with the ITR were predicted by the tfscan tool (Rice *et al.*, 2000). The plugin runs the sequence through the TRANSFAC database, finding possible matches set for a minimum of 7 nucleotides of length and 1 nucleotide maximum mismatch. The ITR secondary structure and status are also used to determine genus-level taxonomic status of novel viruses from the *Parvoviridae* family (Cotmore *et al.*, 2019). The open reading frames (ORFs) were identified with the ORFfinder from NCBI with the following parameters: genetic code: 1; start codon: 'ATG only'; minimal ORF length: 150 nt (<https://www.ncbi.nlm.nih.gov/orffinder/>).

Multiple contigs related to the *Baculoviridae* family were obtained through *de novo* assembly. All contigs and raw reads were mapped to the most similar reference indicated in BLASTn using Geneious mapper with highest sensitivity. From the mapping results, a consensus genome sequence was obtained. Gene predictions were performed using the FgenesV algorithm for circular genomes available at Softberry, which finds similar proteins in public databases and

annotates predicted genes (<http://www.softberry.com/berry.phtml>). Together with an annotation from the most similar reference also performed on Geneious, final ORFs were defined manually.

2.5 Phylogenetic analyses

The assembled viral genomes had their evolutionary history contextualized through similar phylogenetic frameworks. All sequences were aligned with MAFFT v7.505 (Katoh and Standley, 2013) and posteriorly trimmed with TrimAl v.1.4 (Capella-Gutiérrez *et al.*, 2009) with a gap threshold of 0.9, removing sites from alignments composed of 10% or more of gaps. Maximum likelihood trees were inferred with IQ-Tree v.2.1.4 (Minh *et al.*, 2020) with the best fitting model suggested by ModelFinder (Kalyaanamoorthy *et al.*, 2017). Node support was estimated with SH-like approximate likelihood ratio test (SH-aLRT) (Guindon *et al.*, 2010) and ultrafast bootstrap (Minh *et al.*, 2013).

For the *Parvoviridae* genome, the predicted amino acid sequences corresponding to viral structural protein (VP) and non-structural protein (NS1) were identified and had their individual phylogenetic histories inferred with a dataset comprehending the subfamily *Densovirinae*, as described by (Pénzes *et al.*, 2020). For the *Baculoviridae* family, 38 core genes were previously identified through genome sequencing (Javed *et al.*, 2017), as well as an additional set of 25 specific genes in baculoviruses that infect lepidopteran hosts (Garavaglia *et al.*, 2012; van Oers and Vlak, 2007). From those, three were used for phylogenetic analysis. Therefore, nucleotide sequences of *granulin* (*gran*), *late expression factor 8* (*lef-8*) and *late expression factor 9* (*lef-9*) were identified and aligned to previously assembled datasets (Pénzes *et al.*, 2020; Wennmann *et al.*, 2018). Once sequences were aligned and trimmed, the genes were concatenated for phylogenetic inference. Original datasets, alignments and phylogenetic trees can be found in **Supplementary file S1**. Additionally, the alignment of concatenated genes was used to infer a pairwise nucleotide distance comparison using the Kimura two-parameter (K2P) model available on MEGA Software v11.0.13 (Tamura *et al.*, 2021).

3. Results

3.1 Taxonomic composition of the marmoset virome

We characterized the virome from three pools consisting of fecal samples collected from nine free-living hybrids of *C. jacchus* x *C. penicillata*. The library pool of Afetiva (S_Afe) resulted in 1,740,591 raw reads, which after quality control dropped to 751,600 (43,2%) filtered reads (**Table 2**). A large proportion of the filtered reads had no similarity to any sequences available on Genbank (648,172; 86.24%), 103,428 reads (13.76%) were taxonomically assigned

and 690 contigs (192 - 47,091 nt) were *de novo* assembled. The proportion of reads and contigs were 2% and 5% for Viruses, 87% and 89% for Bacteria and 11% and 6% for Eukarya, respectively. Viruses from different taxonomic groups were identified, of which 99% had the closest hits to DNA viruses and 1% to RNA viruses. The viral diversity consisted of several families, including *Siphoviridae*, *Myoviridae* and *Parvoviridae*; unclassified bacteriophages and others with read abundance < 1%. Among the viral families, the *Siphoviridae* family (bacteriophages) was the most abundant, representing 57% of the viruses described.

The library pool of Iguape (S_Igua) resulted in 5,683,230 raw reads, from which 2,771,276 (48,8%) passed quality control. The proportion of assigned reads was <1%, resulting in 228,785 taxonomic reads and 1,798 assembled contigs (129 - 17,496 nt). The proportion of reads and contigs were 7% and 5% for Viruses, 82% and 90% for Bacteria and 11% and 5% for Eukarya, respectively. Viruses from a wide range of taxonomic groups were identified, of which 98% had the closest hits to DNA viruses and 2% to RNA viruses. The viral diversity was composed by the families *Baculoviridae*, *Myoviridae* and *Siphoviridae*; unclassified *Ribovirus* and others with read abundance < 1%. Among the viral families, the *Baculoviridae* (insect viruses) was the most abundant, representing 90% of the viruses described.

The library pool of Igarapé (S_Igar) resulted in 2,516,505 raw reads (1,082,974 filtered reads; 43%). As in the other libraries, a low proportion of reads had similarity with the sequences available in the Genbank database. A total of 126,832 reads were taxonomically assigned and assembled in 3,209 contigs (244 - 6,750 nt). The proportion of reads and contigs were 0.3% and 0.06% for Viruses, 70.7% and 22.9% for Bacteria and 28% and 77% for Eukarya, respectively. Viruses from diverse taxonomic groups were identified, of which 79% had the closest hits to DNA viruses and 21% to RNA viruses. The viral diversity consisted of families *Siphoviridae*, *Myoviridae*, *Podoviridae*, *Autographiviridae*, *Microviridae*, *Parvoviridae*, *Circoviridae*, *Retroviridae* and *Schitoviridae*; unclassified virus such as *Ribovirus* and *CRESS virus* and others with read abundance < 1%. Among the viral families identified, *Siphoviridae* was the most abundant at 16%. A complete table with viral taxonomic assignments can be found in **Supplementary Table S1**.

3.2 Discovery of two novel insect-related viruses

Given trends observed in the abundance of reads and contigs, as well as contig lengths observed for the *Parvoviridae* and *Baculoviridae* families, we were prompted to investigate the occurrence of novel complete genome sequences. In this effort, we characterized two novel viral species and performed comprehensive analyses of their genomes.

3.2.1 *Parvovirus*

In the library pool S_Afe, a contig of 4,573 nt of length and depth of 431x was detected, pertaining to the *Parvoviridae* family with 35,76% identity to nucleotide sequence to *Acheta domesticus mini ambidensovirus* (AdMDV, GenBank: NC_022564.1), detected in common domestic crickets (*Acheta domesticus*) from the United States of America. The final assembly resulted in a genome sequence of 5,309 nt long, with the characteristic *Parvoviridae* 5' and 3' ITRs identified. NCBI ORF finder identified the presence of four main ORFs. Three consecutive 5' ORFs were identified, comprehending 215, 401 and 513 codons, respectively. The second and third ORFs presented long overlapping regions, reflecting distinct reading frames. An additional 3' ORF of opposite-sense presenting 608 codons was also identified resulting in an ambisense genome (ssDNA (+/-)) (**Figure 1A**).

BLASTp tool revealed that the 512-aa protein had 31.8% identity with 89% coverage (E-value 6e-64) with the NS1 and the 607-aa protein had 27.8% identity and 91% coverage (E-value 1e-35) with the VP, both related to AdMDV (Accession No: YP_008658568.1 and YP_008658570.1 respectively). Interestingly, the 400-aa protein was closest to the NS2 protein of *Phoenicurus aureus ambidensovirus* (Accession No: QTE03811), found in feces of Daurian redstart (*Phoenicurus aureus*) in Asia (identity 27.45%, coverage 60% and E-value 3e-09). The smallest protein of 214-aa is a putative NS3 protein from *Densovirinae*, with no similarity to protein sequences available in the database. The NS3 is under the control of the NS3 5' promoter (pNS3) and the promoters for NS2 (pNS2) and NS1 (pNS1) are overlapping and upstream of the NS2 initial codon, while VP is under the control of the pVP promoter. Three core promoter elements (CPEs) - including the B recognition element (BRE), TATA box and initiator element (Inr) - and the downstream promoter element (DPE) are in the pNS3, pNS1/NS2 and pVP promoter regions, whereas the MTE (motif ten element) was not found.

Based on sequence comparison to known densoviruses, two replication initiator motifs called rolling circle replication I and II (RCR I and RCR II) and three helicase motifs were identified in NS1 (**Figure 1A**). The RCR I is located at position 96-103 aa (₉₆HxHxxHDC₁₀₃) containing two conserved histidine residues (H96 and H98) for metal binding. The RCR II is located at position 145-151 aa (₁₄₅Yxxxxxx₁₅₁) and contains a tyrosine residue (Y145) important to participate in the cleavage-linkage reaction. The NS1 also contains the DNA-dependent ATPase/helicase domain with the Walker A, B, and C motifs (aa 369–467). The viral phospholipase A₂ (vPLA₂) was identified in the N-terminal extension of VP protein (aa 6-51) with the calcium binding loop (₁₂GPGN₁₅) and catalytic site (₂₈DxxAxxHDxxY₃₈) found in all

Densovirinae subfamily members. Sequence alignment of the putative NS3 protein revealed four conserved cysteine residues (₁₁₃CxxC55xCxxC₁₇₅) comprising the conserved Zn-finger motif (apoptosis inhibitors) present in all NS3-like arthropod virus proteins. The ORFs and motifs can be found in **Table 3**.

The 5' and 3' telomeric ends of the genome of the new densovirus are, respectively, 145 and 236 nt and form ITRs. The ITRs form imperfect palindromes that conform into secondary structures and present higher GC content (5' ITR: 62.8% and 3' ITR: 52.3%) than the genome as a whole (36.9%). Using Geneious, we identified the minimum free-energy secondary structure of the newly described densovirus (**Figure 1A**). Both ends develop similar trident-shaped secondary structures, and when aligned (data not shown), the shorter 5' ITR is complementary to the central portion of the 3' ITR (trident loop and beginning of stem). This classifies the novel *Densovirinae* herein described as homotelomeric. The most common structure found in other genera of this subfamily are J-shaped telomeres.

The NS1 phylogenetic tree reveals the new densovirus as a sister group to the single representative of the *Miniambidensovirus* genus (AdMDV) with robust node support (SH-aLRT = 100; UFBoot = 99) (**Figure 2**). The described FtDV represents a novel genus, according to the ICTV genera demarcation criteria to the *Parvoviridae* family (*i.e.*, < 40% NS1 amino acid sequence identity) (Walker *et al.*, 2021; Penzes *et al.*, 2020). Different from NS1, the putative VP protein clustered within the recently described sequence (*Muscodensovirus*) with limited branch support (SH-aLRT = 68; UFBoot = 61.3) (**Supplementary Figure S1**). Thus, due to an ambisense genome, format of trident ITRs and the sample type, we named the new virus provisionally as *Fecalis tritonambidensovirus* 1 (FtDV) according to the new binomial nomenclature stipulated by ICTV.

3.2.2 *Baculovirus*

From a *de novo* assembly using reads from the S_Igua pool, a total of 23 *Baculoviridae* contigs were identified, ranging from 230 to 17,496 nt. Following a megablastn analysis, this contig showed 85.7% similarity with *Clostera anastomosis granulovirus B* (ClasGV-B; GenBank: NC_038371.1), a virus of the *Betabaculovirus* genus, specific to the Lepidoptera insect order (moths and butterflies). All 23 contigs were then mapped to this reference genome, covering 88.6% of its extension. After mapping raw reads to this preliminary assembly, genome coverage raised to 100% with 60.8% similarity. The resulting consensus genome sequence had 107,139 nt, a GC content of 34.4% and complementary ends that confirmed circularization.

A total of 153 ORFs with 140 nt or more in length were identified (**Figure 1B**). As previously established, nt 1 was defined based on the adenine of the *granulin* gene, which also

defines orientation of the circular genome. That said, 80 genes were clockwise directed and 73 were counterclockwise. From the 38 core genes, 26 were also identified complete in this novel virus, including two genes necessary for phylogenetic analyses: *orf255* (*lef-8*) and *orf155* (*lef-9*), comprehending 875 and 499 codons, respectively. The other 12 genes were found either incomplete (n=7) or truncated (n=4), and one was not identified at all. Additionally, 20 of the 25 genes found in lepidopteran baculoviruses were present in the genome characterized, including the *granulin* gene (*orf1*; 249 codons), also necessary for phylogeny. Finally, 13 ORFs showed no homology with other genes, being, so far, unique to this virus.

Phylogenetic analysis based on *granulin*, *lef-8* and *lef-9* genes revealed that the novel virus clusters into the *Betabaculovirus* clade, supporting its close relation to ClasGV-B (**Figure 3**). In addition, for lepidopteran-specific baculoviruses, inferences of nucleotide pairwise distances with Kimura two-parameter (K2P) have also been validated for species demarcation: distances up to 0.021 designate viruses belonging to the same species, values up to 0.072 relies on more information and over this threshold indicates viruses of different species (Wennmann *et al.*, 2018). In this case, the genetic distance from the closest reference (ClasGV-B) was calculated as 0.144, confirming that the virus proposed here belongs to a new species, provisionally named *Callithrix fecalis granulovirus* (CafeGV).

3.3 Arthropods signals in metagenomes

In all pools, sequences related to arthropods were obtained. These sequences were often related to groups of hosts known for viruses related to the ones described in this study. In the S_Afe library, where FtDV was found, about 8% (948) of Eukarya reads were mapped to the arthropod phylum, mounted in 25 contigs (375 to 742 nt in length) (**Supplementary Table 2**). The main families found in terms of relative abundance of contigs were 48% *Pyroglyphidae*, 12% *Sarcoptidae* and 4% *Psoroptidae* (Order Sarcoptiformes - mites); 20% of *Tephritidae* (Order Diptera - fruit fly), followed by 4% each of *Pyalidae* (Order Lepidoptera - moths), *Araneidae* (Order Araneae - spiders), *Ixodidae* (Order Ixodida - tick) and *Trombiculidae* (Order Trombidiformes - mites) (**Figure 4**).

In the Iguape pool (S_Igua), where CafeGV was sequenced, about 1% (307) of Eukarya reads were mapped to the arthropod phylum. However, only three contigs were obtained (412, 413 and 1131 nt in length). The arthropod families corresponding to these contigs were 33.3% each for *Calliphoridae*, *Tephritidae* (both Diptera order - fruit fly) and *Daphniidae* (Diplostraca order - water flea).

Although no arthropod-related viruses were found in the Igarape pool (S_Igar), this library had the greatest amount and diversity of arthropod sequences. About 78% (27,315) of Eukarya reads were mapped to the arthropod phylum, mounted in 2,372 contigs (244 to 6,750 nt in length). In all, 21 orders encompassing 96 families of arthropods were sequenced, where the most abundant were orders Sarcoptiformes, mainly the families *Pyroglyphidae* (55%) and *Sarcoptidae* (17,6%), and Diptera, most abundant family *Tephritidae* (17,2%).

4. Discussion

High throughput sequencing is a powerful tool for microbial metagenomic analysis. It has been used to discover a large diversity of novel viruses that could not have been identified by traditional methods (Cosentino *et al.*, 2022; D'arc *et al.*, 2020; Harvey and Holmes, 2022; Shi *et al.*, 2016; Zayed *et al.*, 2022). Following previous virome studies based on non-invasive samples (A Duarte *et al.*, 2019; Bodewes *et al.*, 2014; Wang *et al.*, 2019), we characterized the viral metagenomic of fecal samples of *C. jacchus* x *C. penicillata* hybrids, the first virome analysis from introduced free-living marmosets. Consistent with the observation that it is particularly challenging to identify virus-host associations from fecal samples using HTS (Ge *et al.*, 2012; Waller *et al.*, 2022; Wang *et al.*, 2019), we identified two new likely dietary-related viruses from the *Parvoviridae* and *Baculoviridae* families, both related to insect hosts.

Although the aim in virome studies is sequencing viruses, a large proportion of non-viral genomes were also sequenced in the present study. Previously projects have shown through metagenomics in feces the great diversity of insects consumed by NP, revealing an existing ecological network (Pickett *et al.*, 2012). Silvestre *et al.* (2016) indicated that arthropods constitute about 61% of the diet of the *C. jacchus*. Therefore, it is conceivable to assume that the abundance of insect-related viruses in our samples is a reflection of the marmoset insectivorous diet (Pickett *et al.*, 2012). Noticeably, most of the novel sequences related to insect viruses have low genetic identities with known viruses, indicating that a number of novel viruses exist in the largely uncharacterized virome of Neotropical insect populations (Ali *et al.*, 2021; Maia *et al.*, 2019; Scarpassa *et al.*, 2019). In addition, the virome of marmosets' feces was very similar to that observed in the oral cavity of free-living capuchin monkeys (*Sapajus nigritus*) (Dos Santos *et al.*, 2020). Both primates are considered opportunistic omnivores and feed on a wide variety of food types, including fish, crustaceans, and insects (Galetti and Pedroni, 1994). Insect-related viral families were also sequenced in capuchin monkeys, demonstrating how metagenomics contributes to the characterization of viruses related to the hosts' diets (D'arc *et al.*, 2018; French *et al.*, 2022; Geoghegan *et al.*, 2021).

We report the complete genome sequence of a novel viral species from the *Parvoviridae* family. This family comprises non-enveloped viruses with linear ssDNA genomes (3.7–6.3 Kb) and characteristic terminal hairpins (Cotmore *et al.*, 2014). Although the putative NS3 had no similarity to any Genbank sequence, analysis of the amino acids sequence exhibits a conserved Zn-finger motif from NS3-like of several insect viruses (Krupovic and Koonin, 2014). ITRs are also essential for replication, encapsulation and genome integration (Siegl and Tratschin, 1987; Zádori *et al.*, 2001). Due to the presence of palindromic repeats, the most common structure found in the *Densovirinae* subfamily are J-shaped telomeres, but they can also be found in T-, I-, Y-, U- or simple hairpin-like structures (Laugel *et al.*, 2022). However, the predicted structure of the ITRs for this new virus presented a unique trident-like shape. Following ICTV's genus demarcation (Pénzes *et al.*, 2020), the new species belong to a new genus. Thus, due to the characteristic of the trident shaped ITRs, the ambisense genome architecture and the type of sample, we propose to name it as *Fecalis tritonambidensovirus 1* (FtDV). Interestingly, in the library in which the new FtDV was identified (S_Afe), only Diptera and Lepidoptera orders are known hosts for densoviruses (genus *Protoambidensovirus*) (Pénzes *et al.*, 2020) and both orders were retrieved from our data. However, FtDV is more closely related to an Orthoptera virus which was not found in the library.

We were also able to characterize the complete genome sequence of a novel member of the *Baculoviridae* family. Baculoviruses are insect-specific pathogens, composed of proteic occlusion bodies (OBs) containing enveloped viruses with circular dsDNA genomes, ranging from 80 to 180 kbp (Harrison *et al.*, 2018). They were traditionally classified based on host range, which include orders Lepidoptera, Hymenoptera and Diptera (Herniou *et al.*, 2003), and the OB's composition, which may be polyhedral or ovocylindrical. However, recent phylogenetic approaches demonstrated that phylogenies using highly conserved genes are also a reliable method of classification (Lange *et al.*, 2004; Wennmann *et al.*, 2018). The ICTV guidelines for new virus names indicate a pattern of host specific name, followed by the designation of OB's composition ((Harrison *et al.*, 2018). Since it is not possible to define the direct host of this novel virus and considering we used feces samples of *Callithrix sp.*, we provisionally named it *Callithrix fecalis granulovirus* (CafeGV). The phylogenetic relationship was done based on three core genes, as accepted in the literature ((Lange *et al.*, 2004; Wennmann *et al.*, 2018)). We note that a number of putative core genes are incomplete or absent in the new genome and highlight this new virus may display more genomic plasticity than previously appreciated for the genus. Additional experiments are mandatory to further clarify its genome structure. Finally, in the S_Igua library, where this new *Betaculovirus* was identified, only the Diptera and Diplostraca

(Crustacea) orders were sequenced, while the Lepidoptera order, host of all previously described *Betabaculoviruses*, was not detected (Jehle *et al.*, 2006). However, it is possible that the original host was underrepresented in this sample or could not be assigned to the right taxonomic level.

It is important to highlight that the arthropod-related sequences obtained by HTS in *Callithrix sp.* feces may not necessarily reflect the diversity of arthropods in the diet, since viral metagenomics has a compositional bias toward sequencing viral agents (Kumar *et al.*, 2018). Thus, it becomes difficult to clearly determine whether the likely hosts have been disadvantaged in the HTS or the range of hosts for the mentioned viral groups is wider than previously known. To untangle both scenarios, further investigations into the diversity and evolution of insect viruses are needed.

5. Conclusion

Our study provides, for the first time, an overview of the fecal virome of introduced free-living *Callithrix sp.* hybrids. Through viral metagenomic sequencing, it was possible to characterize two novel dietary-related complete viral genomes, culminating in the description of a new provisional genus for the subfamily *Densovirinae* and a new viral species for the genus *Betabaculovirus*, contributing to the expansion of the families *Parvoviridae* and *Baculoviridae*, respectively. Further studies are needed to improve our knowledge of the virome of the marmoset hybrids and the parental species, *C. jacchus* and *C. penicillata*, allowing us to better understand the circulation of viruses among these animals and their impact on endemic primate species.

Data Availability

Sequences generated in the present work were submitted to Genbank under the accession numbers: ON911537 and OP589943. The metagenomic data set is also available in the NCBI SRA repository under Bioproject PRJNA847605 (SAMN28946322 - SAMN28946324).

Conflict of interest

The authors declare no conflicts of financial or personal interests.

Funding

This work was supported by the *Conselho Nacional de Desenvolvimento Científico e Tecnológico/CNPq* (grant number 313005/2020-6 to AFS) and *Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro/FAPERJ* (grants numbers E26/202.738/2018, E26/210.122/2018, and E26/211.040/2019 to AFS).

Acknowledgements

We thank the field team of the *Associação do Mico-Leão-Dourado* (AMLD) for their help in collecting samples and logistic support in the field. We also thank the members of the LDDV laboratory for their fantastic teamwork.

REFERENCES

- A Duarte, M., F Silva, J.M., R Brito, C., S Teixeira, D., L Melo, F., M Ribeiro, B., Nagata, T., S Campos, F., 2019. Faecal Virome Analysis of Wild Animals from Brazil. *Viruses* 11. <https://doi.org/10.3390/v11090803>.
- Ali, R., Jayaraj, J., Mohammed, A., Chinnaraja, C., Carrington, C.V.F., Severson, D.W., Ramsubhag, A., 2021. Characterization of the virome associated with *Haemagogus* mosquitoes in Trinidad, West Indies. *Sci. Rep.* 11, 16584. <https://doi.org/10.1038/s41598-021-95842-6>.
- Bergner, L.M., Orton, R.J., da Silva Filipe, A., Shaw, A.E., Becker, D.J., Tello, C., Biek, R., Streicker, D.G., 2019. Using noninvasive metagenomics to characterize viral communities from wildlife. *Molecular Ecology Resources*. 19, 128-143. <https://doi.org/10.1111/1755-0998.12946>.
- Bodewes, R., Ruiz-Gonzalez, A., Schapendonk, C.M.E., van den Brand, J.M.A., Osterhaus, A.D.M.E., Smits, S.L., 2014. Viral metagenomic analysis of feces of wild small carnivores. *Virol. J.* 11, 1-13. <https://doi.org/10.1186/1743-422X-11-89>.
- Buchfink, B., Xie, C., Huson, D.H., 2015. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. <https://doi.org/10.1038/nmeth.3176>.
- Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>.
- Carroll, D., Daszak, P., Wolfe, N.D., Gao, G.F., Morel, C.M., Morzaria, S., Pablos-Méndez, A., Tomori, O., Mazet, J.A.K., 2018. The Global Virome Project. *Science*. 359, 872–874. <https://doi.org/10.1126/science.aap7463>.
- Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 34, i884-i890. <https://doi.org/10.1093/bioinformatics/bty560>.
- Cosentino, M.A.C., D’arc, M., Moreira, F.R.R., Cavalcante, L.T. de F., Mouta, R., Coimbra, A., Schiffler, F.B., Miranda, T. dos S., Viana, G.M., Dias, C.A., Souza, A.R., Tavares, M.C.H., Tanuri, A., Soares, M.A., dos Santos, A.F.A., 2022. Discovery of two novel *Torque Teno* viruses in *Callithrix penicillata* provides insights on *Anelloviridae* diversification dynamics. *Front. Microbiol.* <https://doi.org/10.3389/fmicb.2022.1002963>.
- Cotmore, S.F., Agbandje-McKenna, M., Canuti, M., Chiorini, J.A., Eis-Hubinger, A.M., Hughes, J., Mietzsch, M., Modha, S., Ogliastro, M., Pénzes, J.J., Pintel, D.J., Qiu, J., Soderlund-Venermo, M., Tattersall, P., Tijssen, P., Ictv Report Consortium, 2019. ICTV Virus Taxonomy Profile: Parvoviridae. *J. Gen. Virol.* 100, 367–368. <https://doi.org/10.1099/jgv.0.001212>.
- Cotmore, S.F., Agbandje-McKenna, M., Chiorini, J.A., Mukha, D.V., Pintel, D.J., Qiu, J., Soderlund-Venermo, M., Tattersall, P., Tijssen, P., Gatherer, D., Davison, A.J., 2014. The family Parvoviridae. *Archives of Virology*. 159, 1239-1247. <https://doi.org/10.1007/s00705-013-1914-1>.
- D’arc, M., Furtado, C., Siqueira, J.D., Seuánez, H.N., Ayoub, A., Peeters, M., Soares, M.A., 2018. Assessment of the gorilla gut virome in association with natural simian

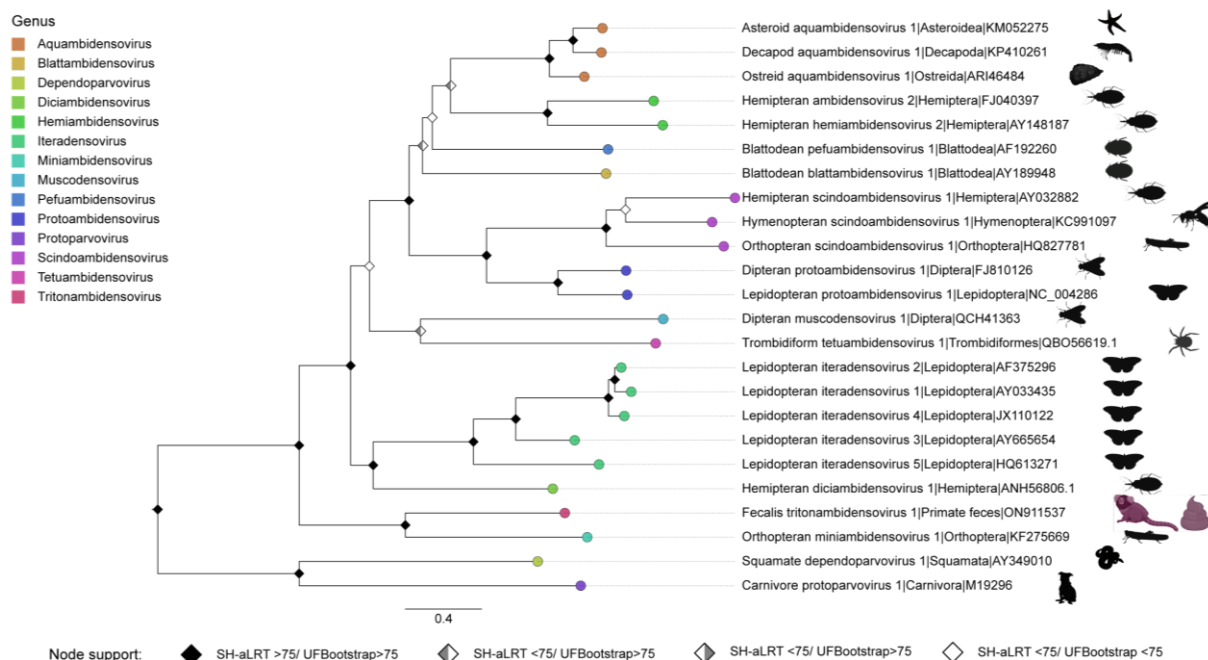
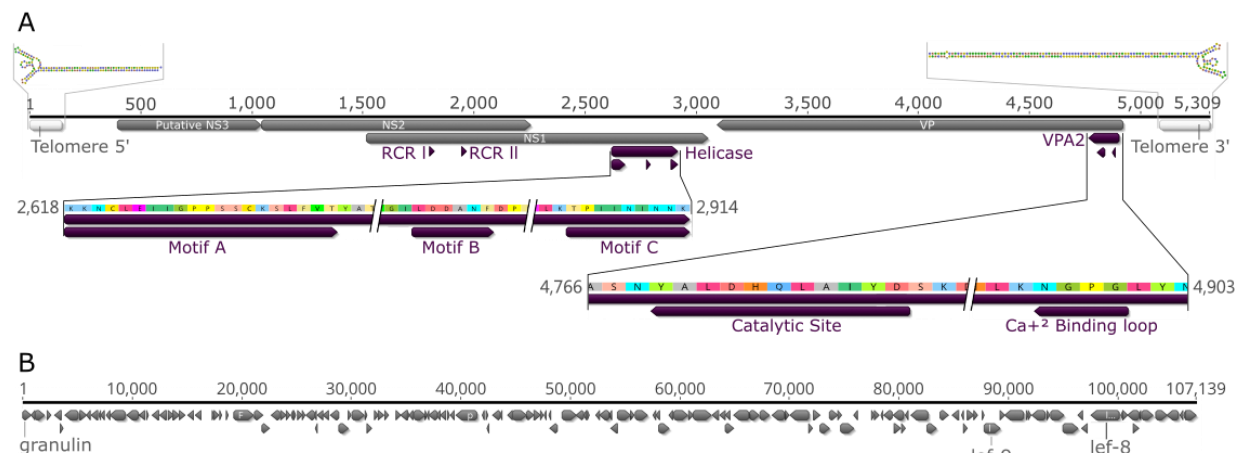
- immunodeficiency virus infection. *Retrovirology* 15, 1-8. <https://doi.org/10.1186/s12977-018-0402-9>.
- D'arc, M., Moreira, F.R.R., Dias, C.A., Souza, A.R., Seuánez, H.N., Soares, M.A., Tavares, M.C.H., Santos, A.F.A., 2020. The characterization of two novel neotropical primate papillomaviruses supports the ancient within-species diversity model. *Virus Evol.* 6, veaa036. <https://doi.org/10.1093/ve/veaa036>.
- Digby, L.J., 1995. Social organization in a wild population of *Callithrix jacchus*: II. Intragroup social behavior. *Primates* 36, 361–375. <https://doi.org/10.1007/BF02382859>.
- Dos Santos, R.N., Campos, F.S., Finoketti, F., Dos Santos, A.C., Campos, A.A.S., Wagner, P.G.C., Roehe, P.M., de Carvalho Ruthner Batista, H.B., Franco, A.C., 2020. Viral diversity in oral cavity from *Sapajus nigritus* by metagenomic analyses. *Braz. J. Microbiol.* 51, 1941–1951. <https://doi.org/10.1007/s42770-020-00350-w>.
- dos Santos Sales, I., Ruiz-Miranda, C.R., de Paula Santos, C., 2010. Helminths found in marmosets (*Callithrix penicillata* and *Callithrix jacchus*) introduced to the region of occurrence of golden lion tamarins (*Leontopithecus rosalia*) in Brazil. *Vet. Parasitol.* 171, 123–129. <https://doi.org/10.1016/j.vetpar.2010.02.044>.
- Dumas, F., Mazzoleni, S., 2017. Neotropical primate evolution and phylogenetic reconstruction using chromosomal data. *Eur. Zool. J.* 84, 1–18. <https://doi.org/10.1080/11250003.2016.1260655>.
- Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. <https://doi.org/10.1093/nar/gkh340>.
- Farouni, R., Djambazian, H., Ferri, L.E., Ragoussis, J., Najafabadi, H.S., 2020. Model-based analysis of sample index hopping reveals its widespread artifacts in multiplexed single-cell RNA-sequencing. *Nat. Commun.* 11, 1-8. <https://doi.org/10.1038/s41467-020-16522-z>.
- French, R.K., Fillion, A., Niebuhr, C.N., Holmes, E.C., 2022. Metatranscriptomic Comparison of Viromes in Endemic and Introduced Passerines in New Zealand. *Viruses*. 14, 1364. <https://doi.org/10.3390/v14071364>.
- Galetti, M., Pedroni, F., 1994. Seasonal diet of capuchin monkeys (*Cebus apella*) in a semideciduous forest in south-east Brazil. *J. Trop. Ecol.* 10, 27–39. <https://doi.org/10.1017/S0266467400007689>.
- Garavaglia, M.J., Miele, S.A.B., Iserte, J.A., Belaich, M.N., Ghiringhelli, P.D., 2012. The *ac53*, *ac78*, *ac101*, and *ac103* genes are newly discovered core genes in the family Baculoviridae. *Journal of Virology*. 86, 12069-12079. <https://doi.org/10.1128/jvi.01873-12>.
- Geoghegan, J.L., Di Giallonardo, F., Wille, M., Ortiz-Baez, A.S., Costa, V.A., Ghaly, T., Mifsud, J.C.O., Turnbull, O.M.H., Bellwood, D.R., Williamson, J.E., Holmes, E.C., 2021. Virome composition in marine fish revealed by meta-transcriptomics. *Virus Evol.* 7, veab005. <https://doi.org/10.1093/ve/veab005>.
- Ge, X., Li, Y., Yang, X., Zhang, H., Zhou, P., Zhang, Y., Shi, Z., 2012. Metagenomic analysis of viruses from bat fecal samples reveals many novel viruses in insectivorous bats in China. *J. Virol.* 86, 4620–4630. <https://doi.org/10.1128/JVI.06671-11>.
- 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. <https://doi.org/10.1093/sysbio/syq010>.
- Harrison, R.L., Herniou, E.A., Jehle, J.A., Theilmann, D.A., Burand, J.P., Becnel, J.J., Krell, P.J., van Oers, M.M., Mowery, J.D., Bauchan, G.R., Ictv Report Consortium, 2018. ICTV Virus Taxonomy Profile: Baculoviridae. *J. Gen. Virol.* 99, 1185–1186. <https://doi.org/10.1099/jgv.0.001107>.
- Harvey, E., Holmes, E.C., 2022. Diversity and evolution of the animal virome. *Nat. Rev. Microbiol.* 20, 321–334. <https://doi.org/10.1038/s41579-021-00665-x>.
- Herniou, E.A., Olszewski, J.A., Cory, J.S., O'Reilly, D.R., 2003. The genome sequence and

- evolution of baculoviruses. *Annu. Rev. Entomol.* 48, 211–234. <https://doi.org/10.1146/annurev.ento.48.091801.112756>.
- IUCN/SSC Primate Specialist Group, 2021. Primates Diversity by Region. http://www.primate-sg.org/primate_diversity_by_region/ (accessed 15 August 2022).
- Javed, M.A., Biswas, S., Willis, L.G., Harris, S., Pritchard, C., van Oers, M.M., Donly, B.C., Erlandson, M.A., Hegedus, D.D., Theilmann, D.A., 2017. Autographa californica multiple nucleopolyhedrovirus AC83 is a infectivity factor (PIF) protein required for occlusion-derived virus (ODV) and budded virus nucleocapsid assembly as well as assembly of the PIF complex in ODV envelopes. *J. Virol.* 91, e02115-16. <https://doi.org/10.1128/JVI.02115-16>.
- Jehle, J.A., Blissard, G.W., Bonning, B.C., Cory, J.S., Herniou, E.A., Rohrmann, G.F., Theilmann, D.A., Thiem, S.M., Vlak, J.M., 2006. On the classification and nomenclature of baculoviruses: a proposal for revision. *Arch. Virol.* 151, 1257–1266. <https://doi.org/10.1007/s00705-006-0763-6>.
- Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. <https://doi.org/10.1038/nmeth.4285>.
- Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. <https://doi.org/10.1093/molbev/mst010>.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>.
- Krupovic, M., Koonin, E.V., 2014. Evolution of eukaryotic single-stranded DNA viruses of the Bidnaviridae family from genes of four other groups of widely different viruses. *Sci. Rep.* 4, 1-9. <https://doi.org/10.1038/srep05347>.
- Lange, M., Wang, H., Zhihong, H., Jehle, J.A., 2004. Towards a molecular identification and classification system of lepidopteran-specific baculoviruses. *Virology.* 325, 36–47. <https://doi.org/10.1016/j.virol.2004.04.023>.
- Laugel, M., Lecomte, E., Ayuso, E., Adjali, O., Mével, M., Penaud-Budloo, M., 2022. The Diversity of Parvovirus Telomeres, in: *Veterinary Medicine and Science*. IntechOpen. <https://doi.org/10.5772/intechopen.102684>.
- Lorenz, R., Bernhart, S.H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., Hofacker, I.L., 2011. ViennaRNA Package 2.0. *Algorithms for molecular biology.* 6, 1-14. <https://doi.org/10.1186/1748-7188-6-26>.
- Maia, L.M.S., Pinto, A.Z. de L., Carvalho, M.S. de, Melo, F.L. de, Ribeiro, B.M., Silhessarenko, R.D., 2019. Novel Viruses in Mosquitoes from Brazilian Pantanal. *Viruses* 11. <https://doi.org/10.3390/v11100957>.
- Malukiewicz, J., Boere, V., de Oliveira, M.A.B., D’arc, M., Ferreira, J.V.A., French, J., Housman, G., de Souza, C.I., Jerusalinsky, L., R de Melo, F., M Valença-Montenegro, M., Moreira, S.B., de Oliveira e Silva, I., Pacheco, F.S., Rogers, J., Pissinatti, A., del Rosario, R.C.H., Ross, C., Ruiz-Miranda, C.R., Pereira, L.C.M., Schiel, N., de Fátima Rodrigues da Silva, F., Souto, A., Šlipogor, V., Tardif, S., 2021. An introduction to the Callithrix genus and overview of recent advances in marmoset research. *ILAR J.* 61, 110–138. <https://doi.org/10.1093/ilar/ilab027>.
- Malukiewicz, J., Boere, V., Fuzessy, L.F., Grativol, A.D., de Oliveira E Silva, I., Pereira, L.C.M., Ruiz-Miranda, C.R., Valença, Y.M., Stone, A.C., 2015. Natural and anthropogenic hybridization in two species of eastern brazilian marmosets (*Callithrix jacchus* and *C.*

- penicillata). PLoS One. 10, e0127268. <https://doi.org/10.1371/journal.pone.0127268>.
- Minh, B.Q., Nguyen, M.A.T., von Haeseler, A., 2013. Ultrafast approximation for phylogenetic bootstrap. Mol. Biol. Evol. 30, 1188–1195. <https://doi.org/10.1093/molbev/mst024>.
- Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37, 1530–1534. <https://doi.org/10.1093/molbev/msaa015>.
- Nurk, S., Meleshko, D., Korobeynikov, A., Pevzner, P.A., 2017. metaSPAdes: a new versatile metagenomic assembler. Genome Research. 27, 824–834. <https://doi.org/10.1101/gr.213959.116>.
- Ondov, B.D., Bergman, N.H., Phillippy, A.M., 2011. Interactive metagenomic visualization in a Web browser. BMC Bioinformatics. 12, 1–10. <https://doi.org/10.1186/1471-2105-12-385>.
- Penzes, J.J., Canuti, M., Söderlund-Venermo, M., Eis-Huebing, A.M., Ogliastro, M., Harrach, B., 2020. Create three new genera and 19 new species (*Piccovirales: Parvoviridae*). https://ictv.global/ictv/proposals/2020.011D.R.Parvoviridae_3ngen_20nsp.zip (accessed 22 September 2022).
- Perelman, P., Johnson, W.E., Roos, C., Seuánez, H.N., Horvath, J.E., Moreira, M.A.M., Kessing, B., Pontius, J., Roelke, M., Rumpler, Y., Schneider, M.P.C., Silva, A., O’Brien, S.J., Pecon-Slatery, J., 2011. A molecular phylogeny of living primates. PLoS Genet. 7, e1001342. <https://doi.org/10.1371/journal.pgen.1001342>.
- Pickett, S.B., Bergey, C.M., Di Fiore, A., 2012. A metagenomic study of primate insect diet diversity. Am. J. Primatol. 74, 622–631. <https://doi.org/10.1002/ajp.22014>.
- Pinheiro, H.L.N., Mendes Pontes, A.R., 2015. Home range, diet, and activity patterns of common marmosets (*Callithrix jacchus*) in very small and isolated fragments of the Atlantic Forest of northeastern Brazil. Int. J. Ecol. Econ. Stat. 2015, 1–13. <https://doi.org/10.1155/2015/685816>.
- Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: the European molecular biology open software suite. Trends in genetics. 16, 276–277. [https://doi.org/10.1016/s0168-9525\(00\)00204-2](https://doi.org/10.1016/s0168-9525(00)00204-2).
- Ross, C.N., Davis, K., Dobek, G., Tardif, S.D., 2012. Aging phenotypes of common marmosets (*Callithrix jacchus*). J. Aging Res. 2012, 567143. <https://doi.org/10.1155/2012/567143>.
- Ruiz-Miranda, C.R., Affonso, A.G., Morais, M.M. de, Verona, C.E., Martins, A., Beck, B.B., 2006. Behavioral and ecological interactions between reintroduced golden lion tamarins (*Leontopithecus rosalia* Linnaeus, 1766) and introduced marmosets (*Callithrix* spp, Linnaeus, 1758) in Brazil’s Atlantic Coast forest fragments. Braz. Arch. Biol. Technol. 49, 99–109. <https://doi.org/10.1590/S1516-89132006000100012>.
- Scarpassa, V.M., Debat, H.J., Alencar, R.B., Saraiva, J.F., Calvo, E., Arcà, B., Ribeiro, J.M.C., 2019. An insight into the sialotranscriptome and virome of Amazonian anophelines. BMC Genomics. 20, 1–13. <https://doi.org/10.1186/s12864-019-5545-0>.
- Shi, M., Lin, X.-D., Tian, J.-H., Chen, L.-J., Chen, X., Li, C.-X., Qin, X.-C., Li, J., Cao, J.-P., Eden, J.-S., Buchmann, J., Wang, W., Xu, J., Holmes, E.C., Zhang, Y.-Z., 2016. Redefining the invertebrate RNA virosphere. Nature 540, 539–543. <https://doi.org/10.1038/nature20167>.
- Siegl, G., Tratschin, J.-D., 1987. Parvoviruses: agents of distinct pathogenic and molecular potential. FEMS Microbiol. Rev. 3, 433–450. <https://doi.org/10.1111/j.1574-6968.1987.tb02479.x>.
- Silvestre, S.M., Dantas, J.O., da Rocha, P.A., Beltrão-Mendes, R., Ferrari, S.F., 2016. Faecal analysis as a complementary approach to the study of feeding ecology in the common marmoset (*Callithrix jacchus* Linnaeus, 1758). Mastozoología neotropical. 23, 185–191.
- Stumpf, R.M., Gomez, A., Amato, K.R., Yeoman, C.J., Polk, J.D., Wilson, B. A., Nelson,

- K.E., White, B.A., Leigh, S.R., 2016. Microbiomes, metagenomics, and primate conservation: New strategies, tools, and applications. *Biol. Conserv.* 199, 56–66. <https://doi.org/10.1016/j.biocon.2016.03.035>.
- Tamura, K., Stecher, G., Kumar, S., 2021. MEGA11: Molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38, 3022–3027. <https://doi.org/10.1093/molbev/msab120>.
- Temmam, S., Davoust, B., Berenger, J.-M., Raoult, D., Desnues, C., 2014. Viral metagenomics on animals as a tool for the detection of zoonoses prior to human infection? *Int. J. Mol. Sci.* 15, 10377–10397. <https://doi.org/10.3390/ijms150610377>.
- van den Brand, J.M.A., van Leeuwen, M., Schapendonk, C.M., Simon, J.H., Haagmans, B.L., Osterhaus, A.D.M., Smits, S.L., 2012. Metagenomic analysis of the viral flora of pine marten and European badger feces. *Journal of Virology.* 86, 2360-2365. <https://doi.org/10.1128/jvi.06373-11>.
- van Oers, M., Vlak, J., 2007. Baculovirus genomics. *Current Drug Targets.* 8, 1051-1068. <https://doi.org/10.2174/138945007782151333>.
- Walker, P.J., Siddell, S.G., Lefkowitz, E.J., Mushegian, A.R., Adriaenssens, E.M., Alfenas-Zerbini, P., Davison, A.J., Dempsey, D.M., Dutilh, B.E., García, M.L., Harrach, B., Harrison, R.L., Curtis Hendrickson, R., Junglen, S., Knowles, N.J., Krupovic, M., Kuhn, J.H., Lambert, A.J., Łobocka, M., Nibert, M.L., Oksanen, H.M., Orton, R.J., Robertson, D.L., Rubino, L., Sabanadzovic, S., Simmonds, P., Smith, D.B., Suzuki, N., Van Dooerslaer, K., Vandamme, A.M., Varsani, A., Zerbini, F.M., 2021. Changes to virus taxonomy and to the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2021). *Archives of Virology.* 166, 2633-2648. <https://doi.org/10.1007/s00705-021-05156-1>.
- Waller, S.J., Lamar, S., Perry, B.J., Grimwood, R.M., Holmes, E.C., Geoghegan, J.L., 2022. Cloacal virome of an ancient host lineage - The tuatara (*Sphenodon punctatus*) - Reveals abundant and diverse diet-related viruses. *Virology.* 575, 43–53. <https://doi.org/10.1016/j.virol.2022.08.012>.
- Wang, Y., Yang, S., Liu, D., Zhou, C., Li, W., Lin, Y., Wang, X., Shen, Q., Wang, H., Li, C., Zong, M., Ding, Y., Song, Q., Deng, X., Qi, D., Zhang, W., Delwart, E., 2019. The fecal virome of red-crowned cranes. *Arch. Virol.* 164, 3–16. <https://doi.org/10.1007/s00705-018-4037-x>.
- Wennmann, J.T., Keilwagen, J., Jehle, J.A., 2018. Baculovirus Kimura two-parameter species demarcation criterion is confirmed by the distances of 38 core gene nucleotide sequences. *J. Gen. Virol.* 99, 1307–1320. <https://doi.org/10.1099/jgv.0.001100>.
- Zádori, Z., Szelei, J., Lacoste, M.C., Li, Y., Gariépy, S., Raymond, P., Allaire, M., Nabi, I.R., Tijssen, P., 2001. A viral phospholipase A2 is required for parvovirus infectivity. *Dev. Cell.* 1, 291–302. [https://doi.org/10.1016/S1534-5807\(01\)00031-4](https://doi.org/10.1016/S1534-5807(01)00031-4).
- Zayed, A.A., Wainaina, J.M., Dominguez-Huerta, G., Pelletier, E., Guo, J., Mohssen, M., Tian, F., Pratama, A.A., Bolduc, B., Zablocki, O., Cronin, D., Solden, L., Delage, E., Alberti, A., Aury, J.-M., Carradec, Q., da Silva, C., Labadie, K., Poulain, J., Ruscheweyh, H.-J., Salazar, G., Shatoff, E., Bundschuh, R., Fredrick, K., Kubatko, L.S., Chaffron, S., Culley, A.I., Sunagawa, S., Kuhn, J.H., Wincker, P., Sullivan, M.B., Acinas, S.G., Babin, M., Bork, P., Boss, E., Bowler, C., Cochrane, G., de Vargas, C., Gorsky, G., Guidi, L., Grimsley, N., Hingamp, P., Iudicone, D., Jaillon, O., Kandels, S., Karp-Boss, L., Karsenti, E., Not, F., Ogata, H., Poulton, N., Pesant, S., Sardet, C., Speich, S., Stemmann, L., Sullivan, M.B., Sunagawa, S., Wincker, P., Coordinators†, T.O., 2022. Cryptic and abundant marine viruses at the evolutionary origins of Earth’s RNA virome. *Science.* 376, 156-162. <https://doi.org/10.1126/science.abm5847>.

Captions and tables



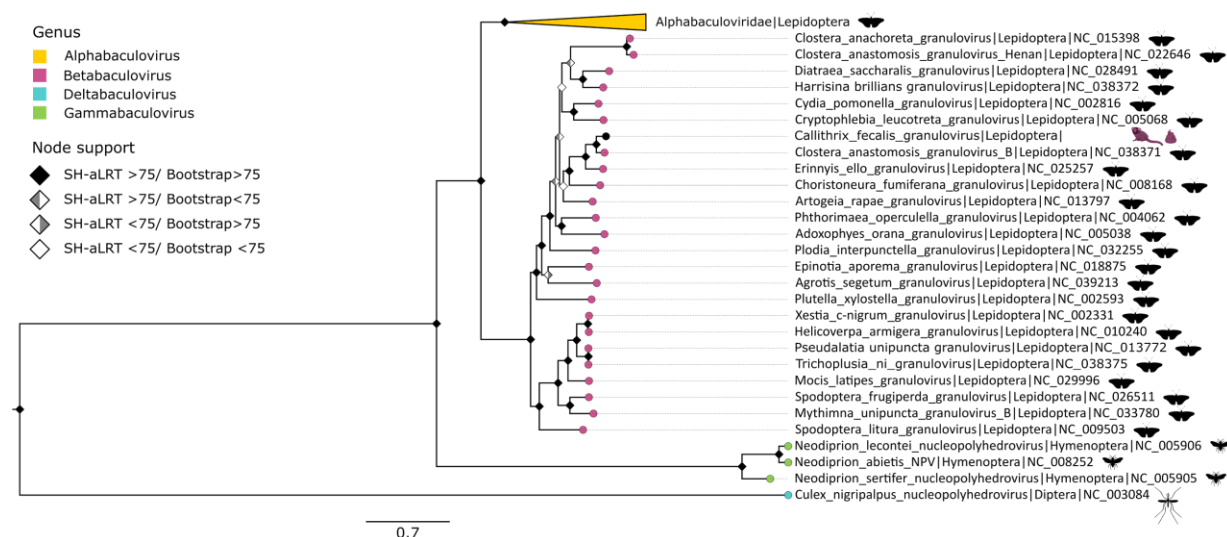


Figure 3: Maximum likelihood phylogeny of the Baculoviridae family. Nucleotide phylogeny of *granulin* (*gran*), *late expression factor 8* (*lef-8*) and *late expression factor 9* (*lef-9*) core genes concatenated, 4,579 bp length under SYM+I+G4 model. The baculovirus identified in the study belongs to the previously described *Betabaculovirus* genus, sister to *Clostera anastomosis granulovirus B*. Circular tip labels are colored according to the *Baculoviridae* genus used in the dataset - *Alphabaculovirus* genus is collapsed and colored in yellow, *Betabaculovirus* in pink, *Deltabaculovirus* in blue and *Gammabaculovirus* in green. *Callithrix fecalis granulovirus* obtained in this study is evidenced by a black tip. Node labels colored in black represent support for SH-aLRT and Bootstrap equal or superior to 75. When only one of the parameters was superior, it is represented in gray, while the other is in white.

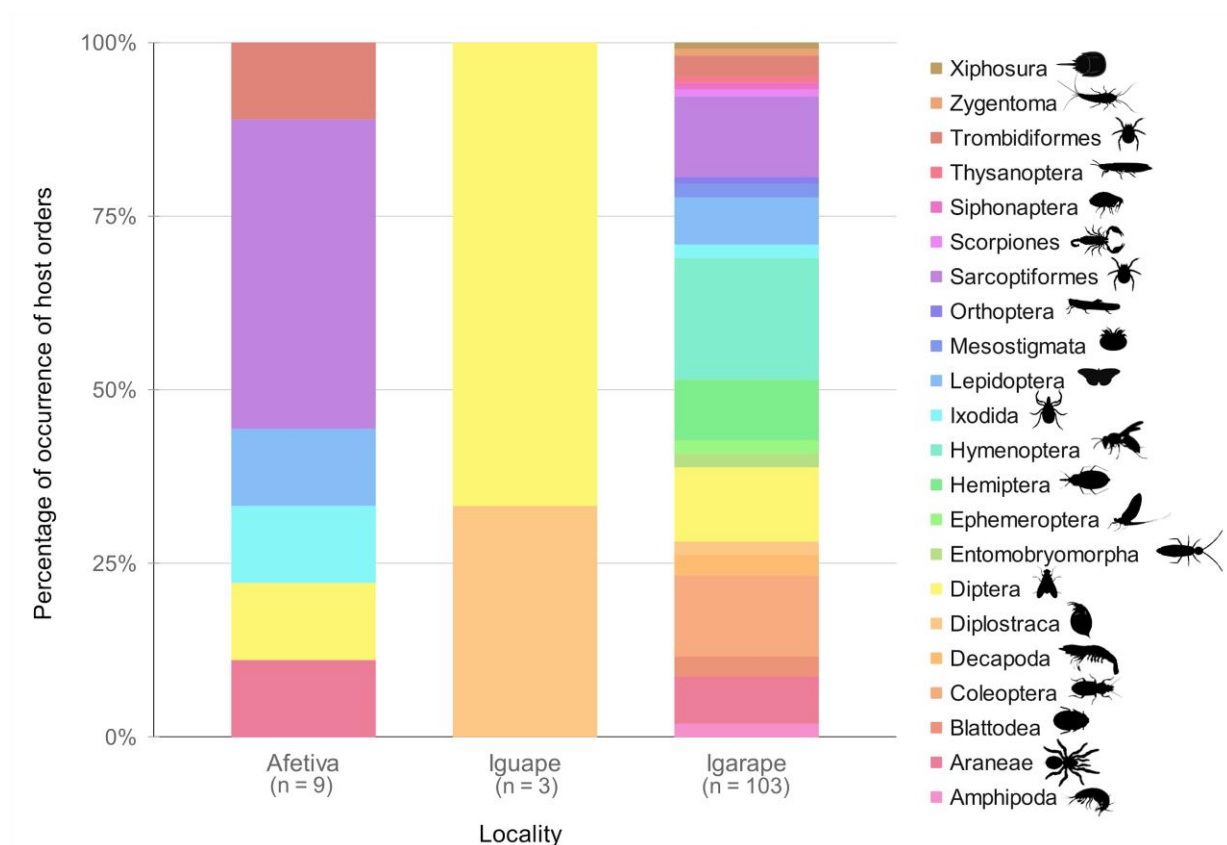


Figure 4: Relative abundance of host order identified by contig taxonomical assignment. Colors and legend on the right refer to the different host orders identified by locality. Columns represent the percentage of host families

within each identified order. A total of 9 families in 6 different orders were identified in Afetiva, 3 families in 2 different orders in Iguape and 103 families in 22 different orders in Igarape.

Table 1: General sampling information.

Species	Location	GPS	Sample Type	Library Code	Number of Individuals
<i>Callithrix sp.</i>	Afetiva, Silva Jardim, RJ	22°37'58.84"S 42°28'1.47"W	Feces	S-Afe	4
<i>Callithrix sp.</i>	Igarapé, Silva Jardim, RJ	22°30'27.41"S 42°18'34.66"W	Feces	S-Igar	2
<i>Callithrix sp.</i>	Iguape, Silva Jardim, RJ	22°30'20.74"S 42°19'24.27"W	Feces	S-Igua	3

Table 2: Overview Virome

	S-Afe	S-Igua	S-Igar
Raw reads	1,740,591	5,683,230	2,516,505
Filtered reads	751,600	2,771,276	1,082,974
G+C content (%)	49%	48%	52%
Mapped reads	103,428	228,785	126,832
Assembled contigs (min - max)	690 (192 - 47,091 nt)	1,798 (129 - 17,496 nt)	3,209 (244 - 6,750 nt)
Reads (%)	89,607 (87%)	186,602 (82%)	90,974 (70.7%)
Bacteria	Contigs (%) (min - max) 615 (89%) (192 - 28,439 nt)	1,613 (90%) (129 - 2,408 nt)	736 (22.9%) (375 - 3,793 nt)
Eukaryota	Reads (%) 11,215 (11%) Contigs (%) (min - max) 39 (6%) (375 - 742 nt)	25,777 (11%) 89 (5%) (132 - 1,580 nt)	35,132 (28%) 2,471 (77%) (244 - 6,750 nt)
Viral	Reads (%) 2,290 (2%) Contigs (%) (min - max) 36 (5%) (378 - 47,091 nt)	14,879 (7%) 84 (5%) (230 - 17,496 nt)	425 (0.3%) 2 (0.06%) (417 - 465 nt)

Table 3: ORFs and motifs

BLASTP analysis							
ORF	Frame	Interval (nt)	Length (nt/aa)	Query Cover	Percent Identity (E-value)	Description (Motifs)	Accession

1	1 (For)	394 - 1,038	(645 / 214)	-	-		Putative NS3 (Zn-Finger)	-
2	3 (For)	1,053 - 2,255	(1,203 / 400)	60%	27.45% (3,00E-09)		NS2	QTE03811.1
3	2 (For)	1,514 - 3,052	(1,539 / 512)	89%	31.84% (6,00E-64)		NS1 (Replication initiation motifs and DNA-dependent ATPase/helicase)	YP_008658568.1
4	1 (Rev)	4,918 - 3,095	(1,824 / 607)	91%	27.88% (1,00E-35)		Structural Protein VP (Viral phospholipase A2)	YP_008658570.1