Viral metagenomics of synanthropic urban bats: a surveillance strategy for uncovering potentially zoonotic viruses

Juliana Amorim Conselheiro[1]; Gisely Toledo Barone[1]; Adriana Araujo Reis-Menezes[1]; Adriana Rückert da Rosa[2]; Débora Cardoso de Oliveira[2]; Filipe Romero Rebello Moreira[3]; Bárbara Aparecida Chaves[3]; Vanderson de Souza Sampaio[3]; Felipe Rocha[4]; Marco Antonio Natal Vigilato[4]; Rodrigo Guerino Stabeli[5,6,7]; Rodrigo Fabiano do Carmo Said[5]; Gabriel Luz Wallau[8,9,10]; Paulo Eduardo Brandão[11]; Anderson Fernandes de Brito[3]

1. Laboratory of Diagnostics of Zoonosis and Vector-Borne Diseases, São Paulo, São Paulo, Brazil
2. Laboratory of Identification and Research of Synanthropic Fauna, São Paulo, São Paulo, Brazil
3. Instituto Todos pela Saúde, São Paulo, São Paulo, Brazil
4. Pan American Center for Foot-and-Mouth Disease and Veterinary Public Health, Pan American Health Organization/World Health Organization (PANAFTOSA/SPV-PAHO/WHO)
5. Pan American Health Organization (PAHO)/World Health Organization (WHO)
6. Ministério da Saúde, Brazil
7. Fundação Oswaldo Cruz, Ribeirão Preto, São Paulo, Brazil
8. Fundação Oswaldo Cruz, Recife, Pernambuco, Brazil
9. Universidade Federal de Santa Maria, Santa Maria, Rio Grande do Sul, Brazil
10. National Reference Center for Tropical Infectious Diseases, Hamburg, Germany
11. University of São Paulo, São Paulo, São Paulo, Brazil

Abstract

Zoonotic spillover events are increasing in frequency, highlighting the urgent need for integrated surveillance systems capable of detecting emerging viruses at the human–animal–environment interface. Bats are natural reservoirs for diverse zoonotic viruses, including coronaviruses, filoviruses, and paramyxoviruses. Here, we present a surveillance framework that leverages Brazil's rabies passive surveillance program to detect bat-borne viruses using metagenomic sequencing. We developed an algorithm to select representative samples from 2,422 bats collected across São Paulo state between October 2023 and February 2024. Paired lung and intestine samples from 150 specimens underwent RNA extraction and nanopore sequencing. A custom bioinformatics pipeline enabled the detection of 66 viral contigs from nine zoonotic-relevant families, including Arenaviridae, Coronaviridae, Filoviridae, and Paramyxoviridae. Notably, we report the first detection of a filovirus in bats in the Americas and characterize the presence of multiple potentially public health–relevant viruses in close contact with humans. These findings reveal previously undetected viral diversity in Brazilian bats and demonstrate how existing animal surveillance initiatives can be leveraged to monitor broader pathogen threats. Our approach provides a scalable, cost-effective model for operationalizing surveillance towards a One Health approach, with potential applications in epidemic and pandemic prevention, particularly in low- and middle-income countries.

One-sentence summary line: The study introduces a viral surveillance method using metagenomics in bats, offering a scalable, cost-effective model for operationalizing One Health surveillance, especially relevant to low- and middle-income countries.

Running head: Viral metagenomics of synanthropic urban bats

Keywords: Metagenomics, bats, surveillance, zoonosis, Nanopore

Viral metagenomics of synanthropic urban bats: a surveillance strategy for uncovering potentially zoonotic viruses

Introduction

Approximately 75% of emerging infectious diseases over recent decades have had zoonotic origins, with the majority involving wildlife reservoirs (Mackenzie et al. 2016). Among these, bats stand out as significant hosts due to their exceptional viral diversity (Letko et al. 2020). To date, over 32,000 viral sequences linked to bats have been deposited in NCBI GenBank, around 88% of which belong to RNA viruses. These include members of families such as *Coronaviridae* (e.g., SARS, MERS), *Rhabdoviridae* (e.g., Rabies virus), and *Paramyxoviridae* (e.g., Hendra and Nipah viruses) (Van Brussel & Holmes 2022).

Bats represent nearly 25% of all known mammalian species and inhabit a wide range of ecosystems globally (Festa et al. 2023). In Brazil, the order Chiroptera includes nine families, 68 genera, and 186 species (Garbino et al. 2024), of which at least 84 have adapted to urban environments (Nunes et al. 2017). In cities such as São Paulo, members of the Phyllostomidae, Molossidae, and Vespertilionidae families are commonly encountered (Almeida et al. 2015). Synanthropic species — those that thrive in close association with human environments — are particularly relevant to public health surveillance. These animals harbor diverse viruses asymptomatically, and their presence in urban settings increases the likelihood of zoonotic transmission. In this context, the human–bat interface represents a dynamic frontier for emerging infectious diseases.

Despite the recognized risks, viral surveillance in bat populations remains limited. Traditional methods such as polymerase chain reaction (PCR), loop-mediated isothermal amplification (LAMP), and enzyme-linked immunosorbent assays (ELISA) are widely used for pathogen detection but are typically restricted to known viruses (Kawasaki et al. 2023). In contrast, metagenomic sequencing offers a powerful alternative for unbiased viral discovery, capable of identifying divergent or novel viruses in a single assay (Miller et al. 2013; Rahimian & Panahi 2024). Although cost and computational demands remain as challenges, the broad-spectrum nature of metagenomic data presents significant advantages.

In Brazil, the Laboratory of Diagnostics of Zoonosis and Vector-Borne Diseases (LabZoo/DVZ/COVISA/SEABEVS/SMS/PMSP) receives approximately 3,000 bats per year through the city's rabies passive surveillance program. Passive surveillance, particularly for wildlife, refers to specimen collection initiated by citizen or health authority reports rather than active field sampling (Minter et al. 2024). While such programs may introduce sampling biases,

they represent cost-effective platforms for detecting circulating or emerging pathogens in large urban areas (Gilbert & Cliffe 2016).

Despite the availability of these specimens, there is currently no integrated strategy to leverage them for broad viral discovery or early zoonotic risk detection. As global health increasingly embraces the One Health paradigm, utilizing urban wildlife surveillance for early viral detection is both a practical and necessary innovation. To address this gap, we developed a novel, cost-effective metagenomic surveillance framework tailored to passive rabies surveillance samples. Our strategy combines an objective sample selection algorithm with SMART-9N/Nanopore sequencing and a robust bioinformatics pipeline, applied to paired lung and intestinal tissues from synanthropic bats collected in São Paulo State, Brazil.

Results

A total of 312 (13%) of all 2,422 bats received at LabZoo for rabies diagnostics through passive surveillance were found eligible for viral metagenomic sequencing, from which 150 specimens with the highest scores (>4 points) were selected for sequencing (6.19% - 150/2,422), totalizing 300 samples (intestines and lungs separately). The samples came from 14 municipalities within the State of São Paulo, Brazil (**Fig. 1A**; **B**), which together account for over 17.3 million inhabitants and have an average population density of approximately 3,220 inhabitants/km^2.

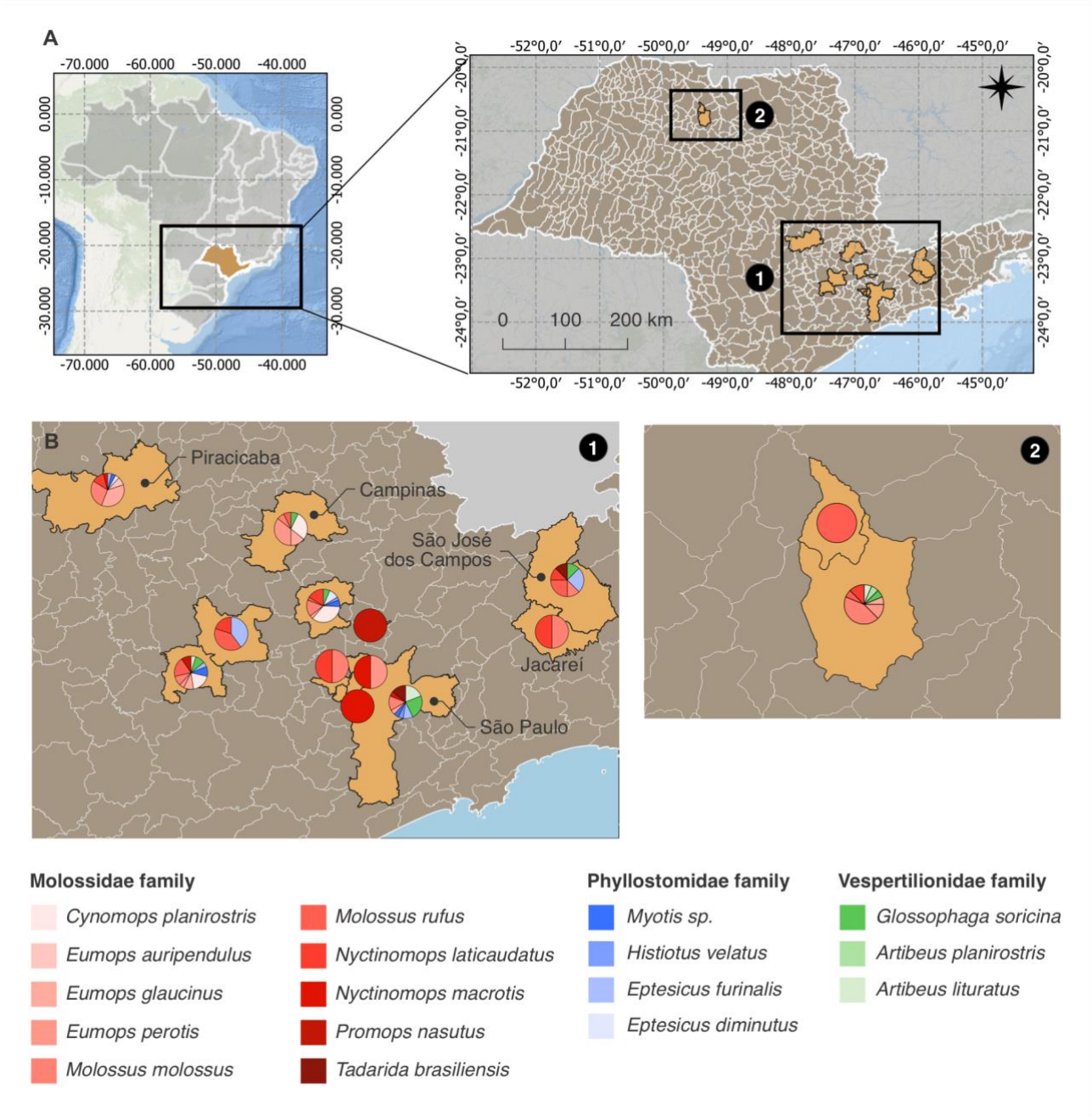


Figure 1. Origin of the 150 bat submitted to viral metagenomics in the State of São Paulo, Brazil. **A.** A total of 14 municipalities within the State of São Paulo were represented. **B.** Colored circles within each municipality represent the bat species.

The bat species belonged to three families: Molossidae (n = 112 specimens), Phyllostomidae (n = 19 specimens), and Vespertillionidae (n = 19 specimens). The greatest diversity was found amongst the Molossidae family, being *Molossus molossus* the most represented species (n = 28) (**Fig. 2A**). As recommended by the rabies surveillance system, bats found fallen on the ground near human dwellings (dead or alive) are constantly reported by local citizens. As such, the majority of the specimens included in this study were found in that context (**Fig. 2B**).

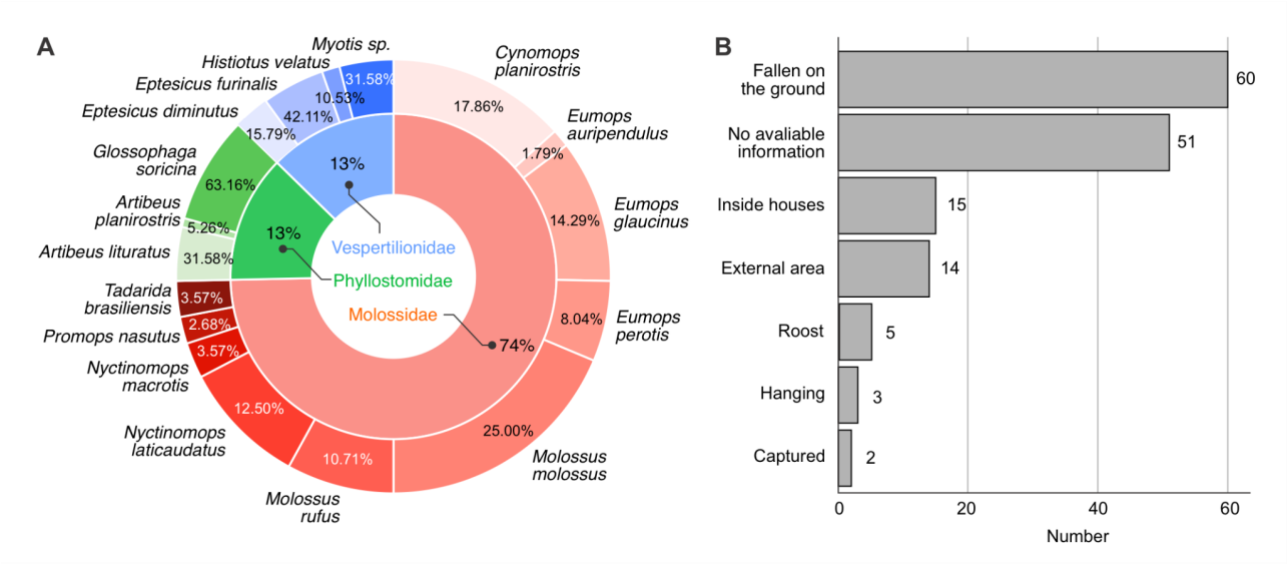


Figure 2. Distribution of bat species and atypical situations of collection. **A.** Distribution of bats species used in viral metagenomics, categorized by family: Molossidae, Phyllostomidae, and Vespertillionidae. **B.** Atypical situations in which bats used in metagenomics were collected.

Metagenomics analysis of bats collected through passive rabies surveillance enables detection of viruses with zoonotic potential

A total of 34.3 million nanopore sequencing reads were generated in this study, from which 88.2% passed quality control **(Table 1**). A total of 589,054 (approximately 1.72%) of these reads were mapped against *de novo* viral assembled contigs (n = 483), spanning 16 distinct viral families. A complete list of these contigs is available on **Supplementary Table 1**. To ensure the reliability of the viral assignments, we applied a series of minimum filters to avoid artifactual detection of viruses, including the establishment of contig length and coverage thresholds, index-hopping filters and secondary similarity searches against a broader database (see Methods). We additionally focused on contigs assigned to viral families known to harbor at least one virus capable of infecting humans. After these filtering steps, 62 contigs remained (**Tables 1** and **2**), 44 assembled from intestines samples and 18 from lung samples, representing 9 distinct viral families (*Arenaviridae*, *Coronaviridae*, *Filoviridae*, *Flaviviridae*, *Paramyxoviridae*, *Phenuiviridae*, *Picornaviridae*, *Rhabdoviridae*, and *Sedoreoviridae*). A total of 9.3% (14/150) bats showed viral hits in at least one of the viscera analyzed, comprehending 8 distinct species (**Table 2**).

136 **Table 1.** Metrics obtained by viral metagenomics of 150 bat from São Paulo State, Brazil.

Viscera	Total raw reads	Total trimmed reads	Total viral reads	Total viral contigs	Filtered viral reads*	Filtered viral contigs*
Intestines	11,471,903	9,679,906	317,926	301	14,901 (4.69%)	44 (14.62%)
Lungs	22,860,735	20,613,332	271,128	182	12.911 (4.76%)	18 (9.89%)
TOTAL	**34,332,638**	**30,293,238**	**589,054**	**483**	**27,812**	**62**

137 *Sequences were filtered as described in the Methods section.

138

139

140

141

142 **Table 2.** Viral hits matching viruses capable of infecting mammals detected in this study.

Viral Family	Best viral hit*	Number of contigs (Intestines)	Number of contigs (Lungs)	Largest contig size (bp)	Reads	Bat ID	Bat species	Dietary habit	Municipality	Collection situation	Reported human contact
Arenaviridae	*Mammarenavirus tacaribeense*	1		1.136	11	3589	*Artibeus lituratus*	Frugivorous	São José do Rio Preto	Unknown	Unknown
Coronaviridae	Myotis ricketti alphacoronavirus Sax-2011		1	1.235	4	3672	*Molossus molossus*	Insectivorous	São Paulo	Fallen on the ground	Unknown
	NL63-related bat coronavirus strain BtKYNL63-9b		1	1.826	6						
Filoviridae	Tai Forest ebolavirus		1	1.191	2	949	*Nyctinomops laticaudatus*	Insectivorous	Jundiaí	Entry	Yes
	Sudan ebolavirus		1	1.150	3						
	Cuevavirus lloviuense		1	1.284	2						
Flaviviridae	*Pegivirus platyrrhini*		2	1.365	4	4449	*Cynomops planirostris*	Insectivorous	Jundiaí	Fallen on the ground	Unknown
	Pegivirus carolliae	1		1.022	2	632	*Eumops glaucinus*	Insectivorous	Piracicaba	External area	Unknown
			1	2.089	4	429	*Eumops glaucinus*	Insectivorous	Piracicaba	Unknown	Unknown
			2	2.898	50	3587	*Glossophaga soricina*	Nectarivorous	São Paulo	Shelter	Unknown
			2	1.683	74	652	*Glossophaga soricina*	Nectarivorous	Sorocaba	Fallen on the ground	Unknown
			2	1.672	144	808	*Glossophaga soricina*	Nectarivorous	Jundiaí	Entry	Yes
			2	1.648	12	3564	*Nyctinomops laticaudatus*	Insectivorous	Jacareí	Unknown	Unknown
	Norway rat hepacivirus 2		1	1.591	822	735	*Nyctinomops laticaudatus*	Insectivorous	Jundiaí	Entry	Yes
Paramyxoviridae	Dolphin morbillivirus		1	1.744	11.784	3538	*Eptesicus furinalis*	Insectivorous	São Paulo	Fallen on the ground	Unknown
Phenuiviridae	Blacklegged tick phlebovirus 1	7		1.102	1.743	3552	*Eumops glaucinus*	Insectivorous	Campinas	Hanging	Unknown
	Blacklegged tick phlebovirus 3	1		628	189						
	Ceraphron bunya-like virus	1		923	507						
	Chagres virus	1		692	287						
	Grand Arbaud virus	1		1.425	599						
	Huangpi Tick Virus 2	5		932	1.513						
	Kabuto mountain virus	3		1.015	1.261						
	Lihan tick virus	1		979	382						
	Mukawa virus	2		767	448						
	Norway phlebovirus 1	3		933	1.400						

	Precarious point virus	1		1.236	416						
	Rukutama virus	1		5.812	1.560						
	Silverwater virus	2		820	909						
	Solenopsis invicta virus 14	2		692	401						
	Toyo virus	4		1.105	1.945						
	Yongjia Tick Virus 1	2		1.260	919						
	Zaliv Terpeniya virus	1		623	356						
Picornaviridae	Shanbavirus A	2		694	13	735	*Nyctinomops laticaudatus*	Insectivorous	Jundiaí	Entry	Yes
Rhabdoviridae	Rhabdoviridae sp.	1		1.137	35	632	*Eumops perotis*	Insectivorous	Campinas	External area	No
Sedoreoviridae	Adult diarrheal rotavirus strain J19	1		657	5	735	*Nyctinomops laticaudatus*	Insectivorous	Jundiaí	Entry	Yes
	Total contigs:	44	18								

143 * Based on primary DIAMOND analyses

In the intestines, we detected 44 viral contigs distributed across six families (**Fig. 3A**). Among these, the family *Phenuiviridae* was the most abundant, represented by 38 high-coverage contigs (average sequencing depth >100×) in a sample from *Eumops glaucinus*. *Flaviviridae* was also detected in another *E. glaucinus* sample. Additional viral families of public health relevance were identified at lower abundances, including *Rhabdoviridae* in *Eumops perotis*, *Arenaviridae* in *Artibeus lituratus*, and *Picornaviridae* and *Sedoreoviridae* in *Nyctinomops laticaudatus* (**Fig. 3A**, **Table 2**).

In the lungs, 18 viral contigs were identified, belonging to four families (**Fig. 3B**). *Paramyxoviridae* was the most abundant, represented by a single high-coverage contig (average sequencing depth >1,500×) in a sample from *Eptesicus furinalis*. *Flaviviridae* was frequently detected across multiple lung samples from diverse host species, including *Cynomops planirostris*, *E. glaucinus*, *Glossophaga soricina*, and *N. laticaudatus*. Of the 18 contigs identified in lung samples, 12 (66.7%) were classified as *Flaviviridae*. Contigs assigned to *Coronaviridae* and *Filoviridae* were also detected at lower abundance in *Molossus molossus* and *N. laticaudatus*, respectively (**Fig. 3B, Table 2**). Notably, *Flaviviridae* was the only viral family detected in both tissue types (**Fig. 3C**).

A comparative analysis of the taxonomic composition by tissue type revealed distinct virome profiles in the intestines and lungs, with *Phenuiviridae* and *Flaviviridae* predominating in each, respectively (**Fig. 3D**). These differences became more pronounced when examining both the number of contigs and read count of each viral family between tissues (**Fig. 3E**). To formally test for compositional differences, we performed a Principal Coordinates Analysis (PCoA) and a PERMANOVA test (**Fig. 3F**). While PCoA revealed some separation between intestine and lung viromes, the PERMANOVA did not yield statistically significant differences ($p > 0.05$), likely due to the limited sample size.

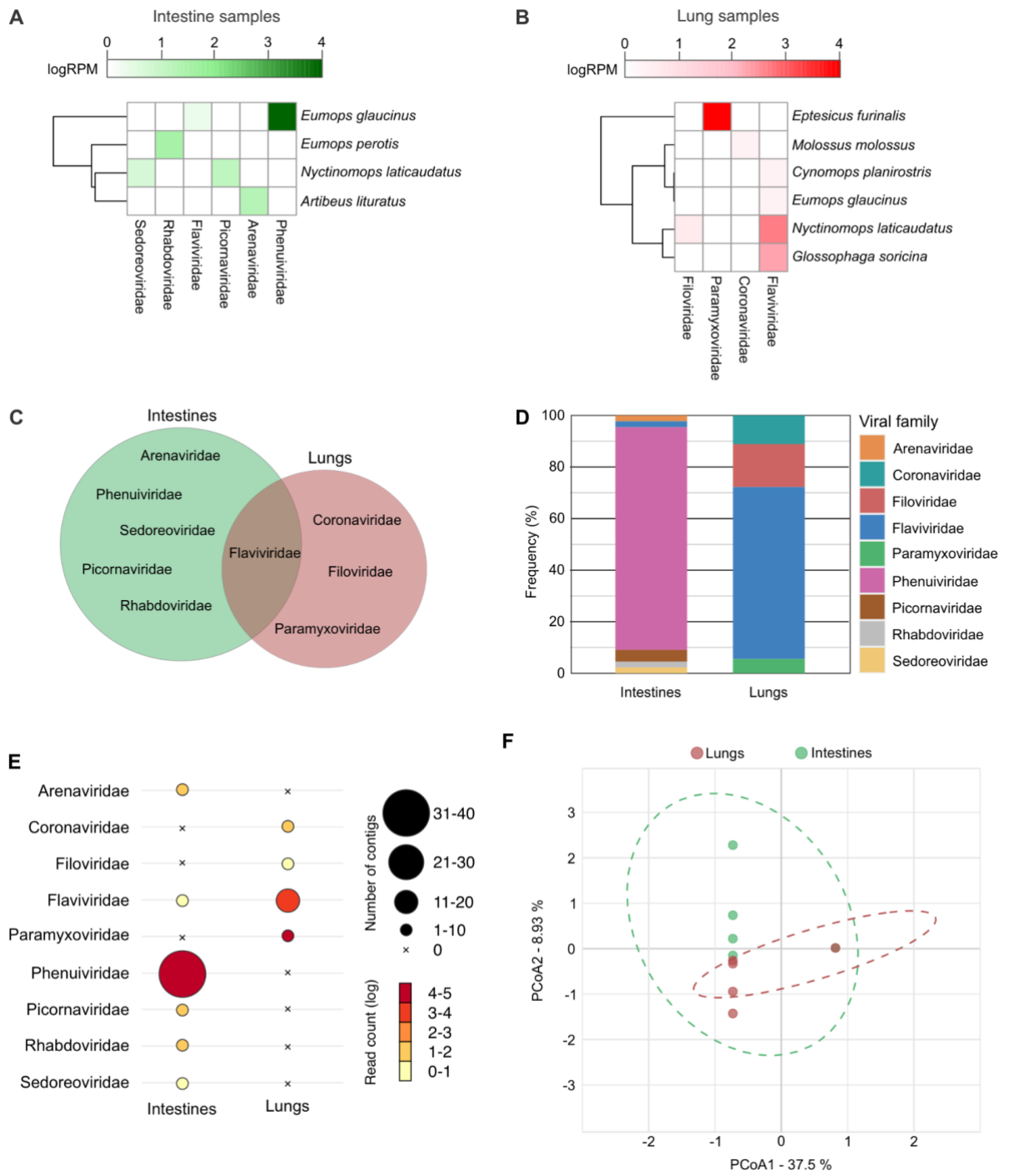
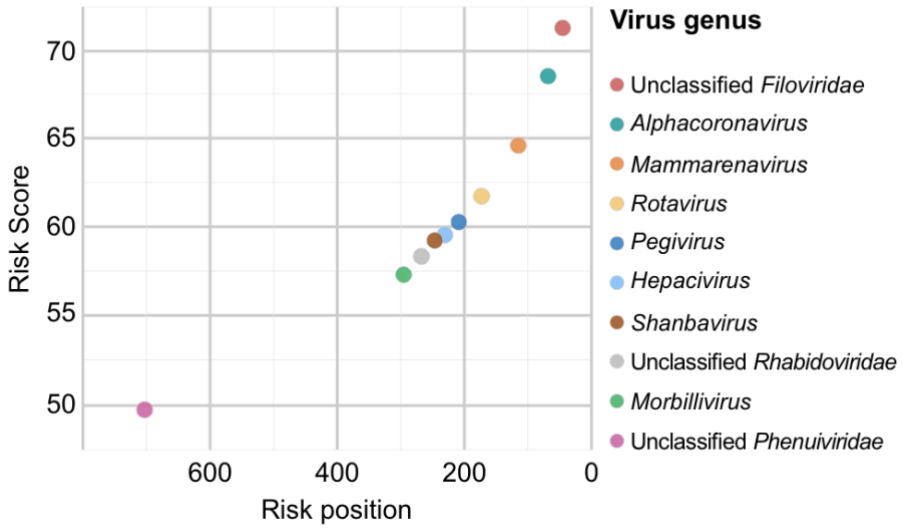


Figure 3. Tissue-specific findings from viral metagenomics analysis of bat specimens. Heatmap of reads mapped against viral sequences per bat species in the intestines (**A**) or lungs (**B**). Values are represented as $\log_{10}$ of mapped reads per million (RPM) and normalized by total reads. (**C**) Venn diagram exhibiting the presence of different viral families between intestines and lungs, revealing that Flaviviridae was the only group present in both tissues. (**D**) Relative frequency of the viral families identified in each tissue. Colors identify distinct viral *taxa*. (**E**) Profile of viral sequences found in intestines and lungs of bat specimens. (**F**) Principal Coordinates Analysis (PCoA) of virome beta diversity across lung and intestinal samples based on Bray-Curtis dissimilarities of viral family abundance profiles. Each point represents a sample, colored by tissue. Although clusters partially overlap, the shift along PCoA1 (37.5% of variance) suggests tissue-specific variation in viral community composition, though this difference was not statistically significant (PERMANOVA, $p > 0.05$).

180

181 ***In silico* assessment supports the spillover potential of detected viruses**

182 As our goal was to use the metagenomics approach as a tool for identifying and monitoring
183 viruses with zoonotic potential in bats, we assessed the spillover or zoonotic re-emergence risk of
184 each viral hit by calculating a risk score through the web tool SpillOver (Grange et al. 2021). This
185 tool calculates the risk factor considering environment, host and the virus itself and provides a risk
186 score and a ranking position. All of the 15 viruses analyzed were considered with moderate to high
187 risk of spillover to humans or zoonotic emergence potential (scores ranging between 50 and 100 –
188 **Fig. 4**). The viral hit with the highest score was the one assigned to the *Filoviridae* family: 71.34
189 (zoonotic risk rank: 16). As per these criteria, a classification in this risk category results in the need
190 for epidemiological and genomic surveillance, *in vitro* studies of virus-host interaction assays as
191 well as environmental surveillance of areas with the presence of natural hosts and humans to track
192 possible changes in viral ecology.



193

194 **Figure 4.** Zoonotic risk scores and ranks of each detected viral *taxa* inferred with the SpillOver web tool. Each point
195 represent a viral *taxa*, either a genus or a sequence yet to be classified at that level (unclassified members of
196 *Phenuiviridae* and *Filoviridae* families). The X-axis shows the spillover risk rank of each viral group, out of 890
197 distinct viruses comprehended in the model. The Y-axis shows the calculated risk score, varying from 0 to 150. For
198 reference, the highest ranking virus in the Spillover model is SARS-CoV-2, with a risk score of 97
199 (https://spillover.global/ranking-comparison/, last accessed June 19th, 2025).

200

201 **Phylogenetic analyses clarify the evolutionary placement of novel viral sequences**

202 To contextualize the novel sequences identified from viral families of particular zoonotic
203 relevance, we performed maximum-likelihood (ML) phylogenetic reconstructions using
204 comprehensive reference datasets. Overall, these analyses were consistent with the known
205 evolutionary history of each viral family and aligned with preliminary classifications based on

206 sequence similarity searches. These reconstructions provided refined insights into the evolutionary
207 placement of the novel sequences relative to the closest available references (**Fig. 5**). All multiple
208 sequence alignments and phylogenetic trees generated in this study are provided in **Supplementary**
209 **File 1**.

210 The *Arenaviridae* nucleoprotein sequence was preliminarily identified as *Mammarenavirus*
211 *tacaribense* based on BLASTp results (97.13% identity, 100% query coverage, e-value: 6e-172).
212 This classification was corroborated by ML phylogenetic analysis (**Fig. 5A**), where the sequence
213 clustered within a clade of previously characterized *M. tacaribense* isolates (SH-aLRT = 83.7)
214 collected between 1956 and 2012 in Brazil, Trinidad and Tobago, and the USA.

215 The *Coronaviridae* Spike protein sequence showed highest similarity to *Tadarida*
216 *brasiliensis* alphacoronavirus 1 (BLASTp results: 73.65% identity, 94% query coverage, e-value:
217 2e-140). Phylogenetic analysis supported this genus-level classification: the novel sequence formed
218 a sister clade (SH-aLRT = 100) to a group of sequences derived from *Eptesicus*, *Pipistrellus*, and
219 *Tadarida* bats, all falling within the *Alphacoronavirus* genus (**Fig. 5B**). These reference sequences
220 originated from bats sampled between 2021 and 2022 in Argentina, Pakistan, the USA, and South
221 Korea.

222 The *Filoviridae* RNA-dependent RNA polymerase (RdRp) sequence (1,014 bp) had as its
223 closest match the homologous protein from *Orthoebolavirus restonense* (BLASTp results: 78.40%
224 identity, 100% query coverage, e-value: 0). However, ML phylogenetic inference indicated that the
225 novel sequence is sister to a clade comprising filoviruses recently characterized in bats from China
226 (NCBI accession numbers: KP233864, KX371873; SH-aLRT = 95.3), with this entire clade
227 positioned as sister to the *Orthoebolavirus* genus (**Fig. 5C**). To further confirm the assignment of
228 this virus, we performed protein structural analysis with HHpred, which confirms our sequence
229 harbors a profile compatible with the RdRp of Ebola virus (hit: 9IP3_A; probability: 100%; e-value:
230 4.9e-71).

231 We also investigated two additional *Filoviridae* contigs in the same sample. One contig (303
232 bp) contains an open reading frame (ORF) encoding a protein with high similarity to the
233 nucleoprotein of *Orthoebolavirus restonense* (BLASTp results: 44.94% identity, 88% query
234 coverage, e-value: 2e-20). This identification was supported both by HHpred structural homology
235 analysis (hit: 5W2B_A, probability: 100%; e-value: 2.5e-52) and by phylogenetic inference, which
236 placed this new nucleoprotein sequence as sister to sequences in the *Orthoebolavirus* clade (SH-
237 aLRT = 93.4; **Supplementary File 1**). The second contig (293 bp) harbored shorter ORFs (~50
238 amino acids) encoding proteins most similar to the glycoprotein (GP2) of distinct filoviruses,
239 including *Dianlovirus menglaense* (BLASTp results: 64.86% identity, 66% query coverage, e-
240 value: 1e-8) and *Orthoebolavirus sudanense* (57.50% identity, 78% query coverage, e-value:

241 0.003). These identifications were further supported by HHpred analyses (hits: 2LCZ_A and
242 2EBO_A; probability >99%, e-value <1e-10), confirming structural features consistent with
243 filoviral glycoproteins.

244 As the formal classification of novel filoviruses requires complete genome sequences, the
245 precise taxonomic status of this virus remains unresolved. Nonetheless, to our knowledge, this
246 represents the first detection of a filovirus in bats from the Americas.

247 The *Paramyxoviridae* matrix protein sequence was most similar to that of *Morbillivirus*
248 *myotis* (BLASTp results: 100% identity, 96% query coverage, e-value: 0). This identification was
249 further supported by phylogenetic analysis, where the sequence clustered within a clade of *M.*
250 *myotis* sequences (SH-aLRT = 87.8) previously sampled in Brazil in 2013 and 2015 (**Fig. 5D**).

251 The *Phenuiviridae* RNA-dependent RNA polymerase (RdRp) protein sequence showed
252 highest similarity to *Uukuvirus rukutamaense* (BLASTp results: 35.72% identity, 97% query
253 coverage, e-value: 0). However, in the maximum-likelihood (ML) phylogenetic tree, the sequence
254 clustered with *Drosophila* Sighthill phlebovirus (SH-aLRT = 96.7), an unclassified member of the
255 family (**Fig. 5E**). Together with Ips phenui-like virus 2, another unclassified virus, these sequences
256 formed a clade that is sister to the genus *Ixovirus*, whose members are known to exclusively infect
257 ticks. Notably, the patristic distances observed among these sequences are comparable to those
258 separating established genera, suggesting they may represent novel clades yet to be formally
259 classified at the genus level.

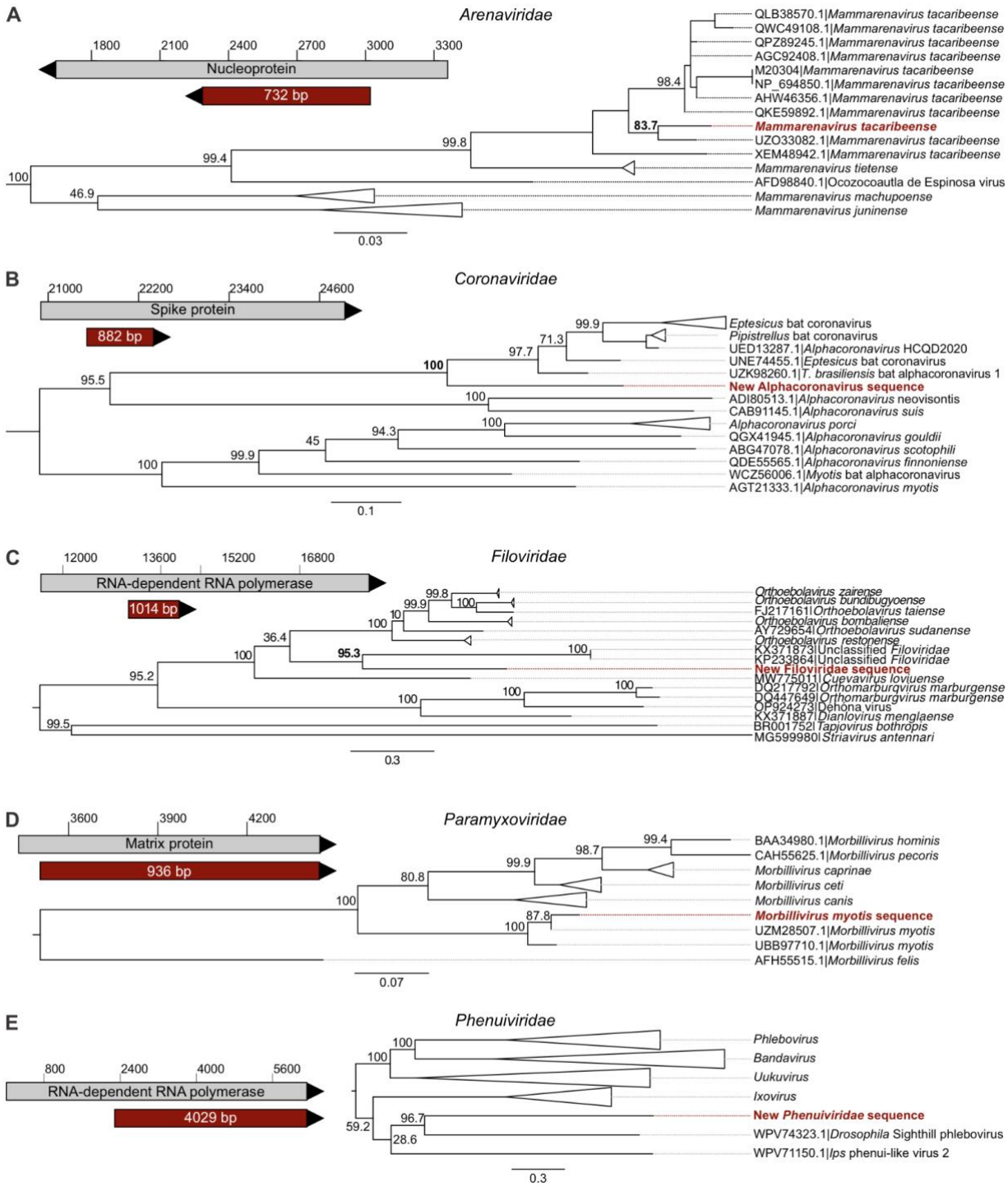


Figure 5. Maximum-likelihood phylogenetic reconstruction of amino acid sequences derived from contigs assigned to selected viral families of zoonotic relevance. Here, only the coding region of the contigs were used for phylogenetic analyses. (**A**) Phylogenetic tree of the nucleoprotein of members of the *Arenaviridae* family, focused on a subtree from the genus *Mammarevirus*. The new sequence clusters within the diversity of the species *Mammarenavirus tacaribeense* with high support value (SH-aLRT = 83.7). The inset shows the genomic context of the contig, mapped to a *Tacaribe virus* sequence (segment S, NC_004293). (**B**) Phylogenetic tree of the Spike protein of the family *Coronaviridae*, exhibiting a part of the *Alphacoronavirus* genus subtree. The new sequence cluster near a clade comprehending bat alphacoronaviruses sampled across the globe (SH-aLRT = 100). The inset shows the contig mapped to a *Eptesicus* bat coronavirus sequence (OL415262). (**C**) Phylogeny of the RdRp of the family *Filoviridae*, exhibiting the phylogenetic

placement of the new sequence as an external group to the *Orthoebolavirus* genus (SH-aLRT = 93.4). The inset shows the contig mapped to a *Reston ebolavirus* sequence (AY769362). (**D**) Phylogenetic tree of the matrix protein of the *Paramyxoviridae* family, exhibiting only the subtree of the *Morbillivirus* genus. The new sequence clusters within the diversity of the species *Morbillivirus myotis* (SH-aLRT = 87.8). The inset shows the contig mapped to a *Myotis* bat morbillivirus sequence (MW557651). (**E**) Phylogenetic tree of the RdRp sequences of the *Phenuiviridae* family, revealing that the new sequence clusters with the *Drosophila* Sighthill phlebovirus, near the clade for the genus *Ixovirus*. The inset shows the contig mapped to a *Rukutama virus* sequence (segment L, KF892052). Tip labels in red font indicate the sequences generated in this study. Scale bars represent amino acid substitutions per site. Collapsed clades indicate sequences belonging to the same taxa at the genus or species level. Complete phylogenetic trees are available in **Supplementary File 1**.

Discussion

Emerging infectious diseases with zoonotic origins have repeatedly demonstrated the need for scalable, pre-emptive viral surveillance in wildlife, especially in synanthropic species that frequently interface with human environments. In this study, we present a metagenomics-based surveillance framework applied to bats sampled through Brazil's passive rabies monitoring system. This approach leverages existing infrastructure to cost-effectively detect viruses with zoonotic potential, including families of high public health concern such as *Filoviridae*, *Paramyxoviridae*, and *Coronaviridae*. Notably, we report what is, to our knowledge, the first detection of a filovirus in bats in the Americas.

Because bats are legally protected in many countries and are difficult to sample longitudinally in the wild, passive surveillance offers an ethical and logistically feasible alternative for accessing specimens (Minter et al. 2024). Although its effectiveness can vary with disease dynamics and public reporting patterns, our findings support its value for virus detection and discovery. At LabZoo, more than 3,000 bats are submitted annually for rabies diagnostics, yet these specimens are typically not screened for other pathogens. Our metagenomics pipeline demonstrates that this underutilized resource can be mined for broad viral surveillance.

We found a distinct separation of viral families between intestinal and pulmonary tissues. The aggregated intestinal virome predominantly harbored *Phenuiviridae*, a diverse group with members capable of infecting arthropods, plants, and vertebrates, including Rift Valley fever virus and Severe Fever with Thrombocytopenia Syndrome virus (Sasaya et al. 2023), whereas the lungs featured more vertebrate-related viruses, including *Arenaviridae* (genus *Mammarenavirus*), *Paramyxoviridae* (genus *Morbillivirus*), *Coronaviridae* (genus *Alphacoronavirus*), and, at a higher frequency, *Flaviviridae* (genus *Pegivirus*). These patterns suggest that tissue tropism, dietary exposure, and possibly arthropod infestation contribute to virome composition (Van Brussel & Holmes 2022; Bazzoni et al. 2024). Although our PERMANOVA did not reach statistical significance, the taxonomic patterns and PCoA clustering indicate biologically meaningful

differences. These findings support incorporating multi-tissue sampling into future viral metagenomics workflows, in agreement with previous recommendations (Wallau et al. 2023).

The presence of tick-associated viruses in intestines, particularly members of *Phenuiviridae*, may reflect ectoparasite ingestion or infestation rather than active infection, consistent with reports of tick-borne viruses associated with bats (Xu et al. 2022; Luz et al. 2016). This raises intriguing ecological questions about the role of ectoparasites in bat viral ecology that merit further investigation. The detection of *Flaviviridae* sequences across multiple samples is also consistent with the literature, which indicates that bats are natural hosts frequently infected with members of the genera *Hepacivirus* and *Pegivirus*, typically without apparent disease (Quan et al. 2013).

Among the vertebrate-associated viruses, we recovered a near-complete genome of a *Myotis* bat morbillivirus, previously identified in Brazil and shown to recognize human cell receptors *in vitro* (Wells et al. 2022; Ikegame et al. 2023). Given the known capacity of paramyxoviruses to cross species barriers (Branda et al. 2024), this detection underscores the need for continuous surveillance. We also detected a sequence closely related to *Mammarenavirus tacaribeense*, which has been reported in frugivorous bats and ticks (Bentim Góes et al. 2022; Sayler et al. 2014), suggesting possible multi-host dynamics. Likewise, a partial sequence of a rotavirus genetically similar to the human strain J19 was found, raising questions about reverse zoonosis or shared environmental sources. The detection of alphacoronaviruses such as *Tadarida brasiliensis* alphacoronavirus 1 aligns with previous findings (Cerri et al. 2023; Schaeffer et al. 2022) and adds to the growing recognition of the diversity of bat-borne coronaviruses in South America.

Perhaps the most striking finding was the identification of contigs homologous to the RNA-dependent RNA polymerase (RdRp), nucleoprotein, and glycoprotein of filoviruses, supported by both phylogenetic and protein structure analyses. Although the sequences were consistently divergent, evolutionary analyses placed the novel virus in a clade with filoviruses previously detected in bats from China (Yang et al. 2017). This group forms a sister lineage to the genus *Orthoebolavirus* and shares structural features with its members. Despite only partial genomic recovery, the assembled fragments mapped to functionally important regions of multiple filoviral proteins, supporting their potential biological relevance. This constitutes the first report of a filovirus in bats from the Americas and underscores the importance of global surveillance beyond regions historically associated with filovirus outbreaks.

Given that multiple viruses were detected in bats collected in atypical situations *(i.e.* in close contact with humans or domestic animals), we were prompted to further investigate their zoonotic potential. Using the SpillOver tool (Grange et al. 2021), we found that all 15 detected viral *taxa* were categorized as moderate to high risk for zoonotic emergence. The top-scoring virus — our candidate filovirus — ranked 16th among the 890 viruses included in the model. This *in silico* risk

evaluation helps prioritize viruses for further characterization and offers a model for integrating genomic surveillance data with threat prediction tools.

Our study is limited by potential nucleic acid degradation due to variable specimen quality, the lack of PCR or culture-based confirmation, and incomplete genome assemblies for many contigs. However, the combination of sequence homology, structural modeling, and phylogenetic inferences provides strong support for our findings. Going forward, integrating targeted PCR assays, virus isolation, and host-receptor interaction studies will be critical to characterize the biological properties of these viruses. We also recommend automation of the decision-tree selection algorithm and integration of ecological metadata to optimize the detection of viruses with zoonotic potential.

Conclusion

Our findings demonstrate that viral metagenomics applied to bat specimens from a rabies passive surveillance system can reveal viruses of zoonotic relevance, including novel or divergent lineages. This approach provides a scalable and cost-effective model for operationalizing One Health surveillance, with potential applications in pandemic prevention, particularly in low- and middle-income countries. Findings like the ones reported in this study are relevant for triggering active surveillance of novel viruses, especially when suspected or confirmed human cases of new zoonotic diseases are reported. As ecological disruption continues to drive cross-species transmission events, strategies like the one proposed here will be critical for the early detection of zoonotic viruses with epidemic and pandemic potential.

Methods

Sample collection and selection protocol

Bats routinely submitted to LabZoo as part of the rabies passive surveillance program were used in this study. The samples originated from 22 municipalities of São Paulo State, Brazil, between October 2023 and February 2024. All bat specimens were analysed and tested for rabies diagnosis, following gold standard procedure, applying direct immunofluorescence assay. Additionally, lung and intestine samples from rabies-negative bats were collected and stored at -80 °C until use. All specimen processing steps were performed in a Class II Biological Safety Cabinet. A decision tree was elaborated to guide selection of samples (**Fig. 6**; see **Supplementary Methods 1** and **Supplementary Table 2**), with specimens scoring between 5 and 7 deemed eligible for viral metagenomic sequencing.

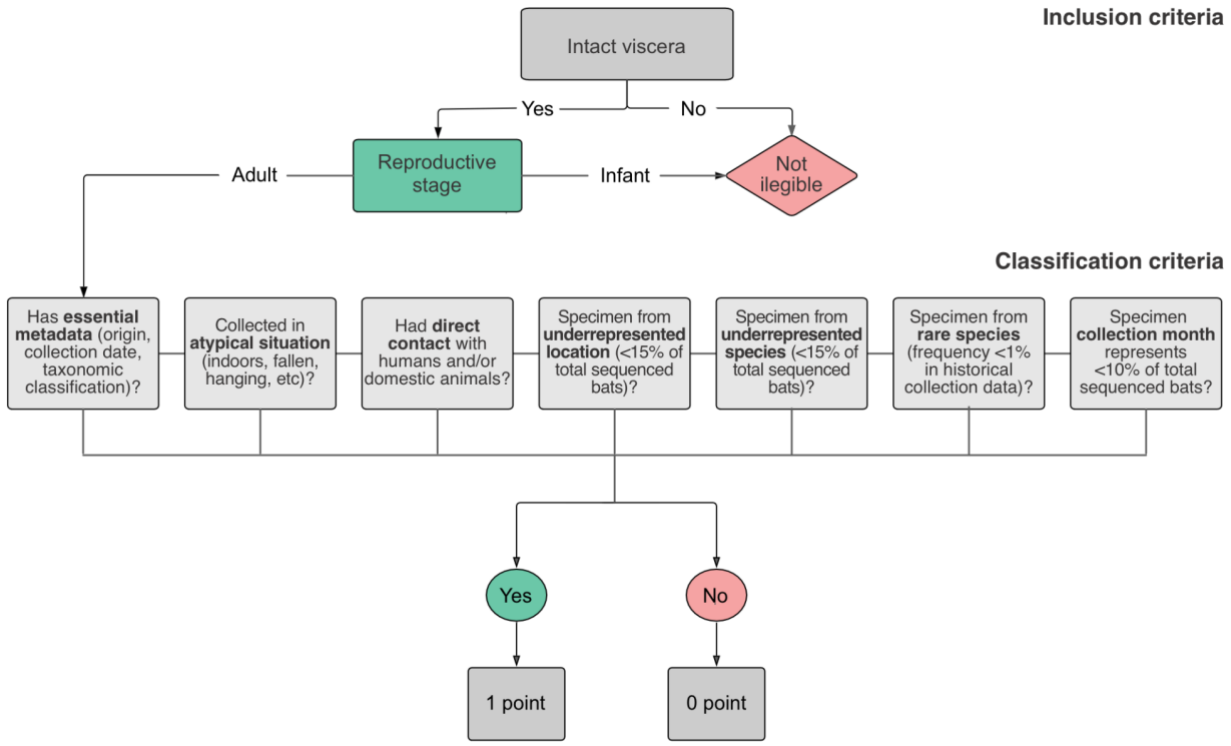


Figure 6. Decision tree used to guide the selection of bats for viral metagenomics. Viscera with vivid red color and visually integral texture were considered intact. Thresholds for municipality representation (<15%), species representation (<5%), rare species (<1%), and collection month (<10%) were calculated using the empirical distribution of previously sequenced samples. Each parameter scores one point whenever its threshold is met. This scoring system was designed to minimize bias and ensure a balanced representation of bats across space, time, and species. See **Supplementary Table 2**.

Nucleic acid isolation

A total of 100uL of Phosphate-Buffered Saline pH 7.4 (PBS) along with 4 glass beads of 1mm diameter were added to 10mg of intestine and/or lung tissue of each bat and then submitted to tissue homogenization in an L-Beader 24 (Loccus, Cotia, SP, Brazil). Next, 500µL of PBS were added to each sample and the tubes were centrifuged at 16,000 x *g* for 5 minutes at 4°C. The supernatants were filtered through 0.45µM syringe filters (Sigma-Aldrich, St. Louis, MO, USA) and submitted to nucleic acid isolation using the Bio Gene Viral DNA/RNA extraction kit (Quibasa, Belo Horizonte, MG, Brazil), following the manufacturer's instructions. The samples were also submitted to DNAse treatment using TURBO DNAse 2U/µL (Invitrogen, Carlsbad, CA, USA), followed by DNAse cleanup using the RNA Clean & Concentrator-5 kit (Zymo Research, Tustin, CA, USA) as described by each manufacturer’s protocol.

398 **Nanopore sequencing**

399 The synthesis of cDNA and PCR reaction were performed using the SMART-9N protocol,
400 as described elsewhere (Claro et al. 2021). PCR products were purified using a 1:1 ratio of AMPure
401 XP beads (Beckman Coulter, High Wycombe, UK), quantified with the fluorometer Qubit 3.0 (Life
402 Technologies, Carlsbad, CA, USA), and normalized to obtain 50ng per sample. Library preparation
403 was carried out using the Ultra II End Repair/dA-tailing Module (NEB) and Ultra Ligation Module
404 (NEB). Barcoding was performed using the EXP-NBD104 and EXP-NBD114 kits along with the
405 Ligation Sequencing kit SQK-LSK109 (ONT, Oxford, UK). Next generation sequencing was
406 carried out in the MinION MK1C (ONT), by applying 50ng of the final libraries onto 14 R9.4.1
407 flow cells (ONT). A total of 24 samples were sequenced per individual flow cell. Sequencing was
408 performed for 72 hours or until complete pore depletion.

409

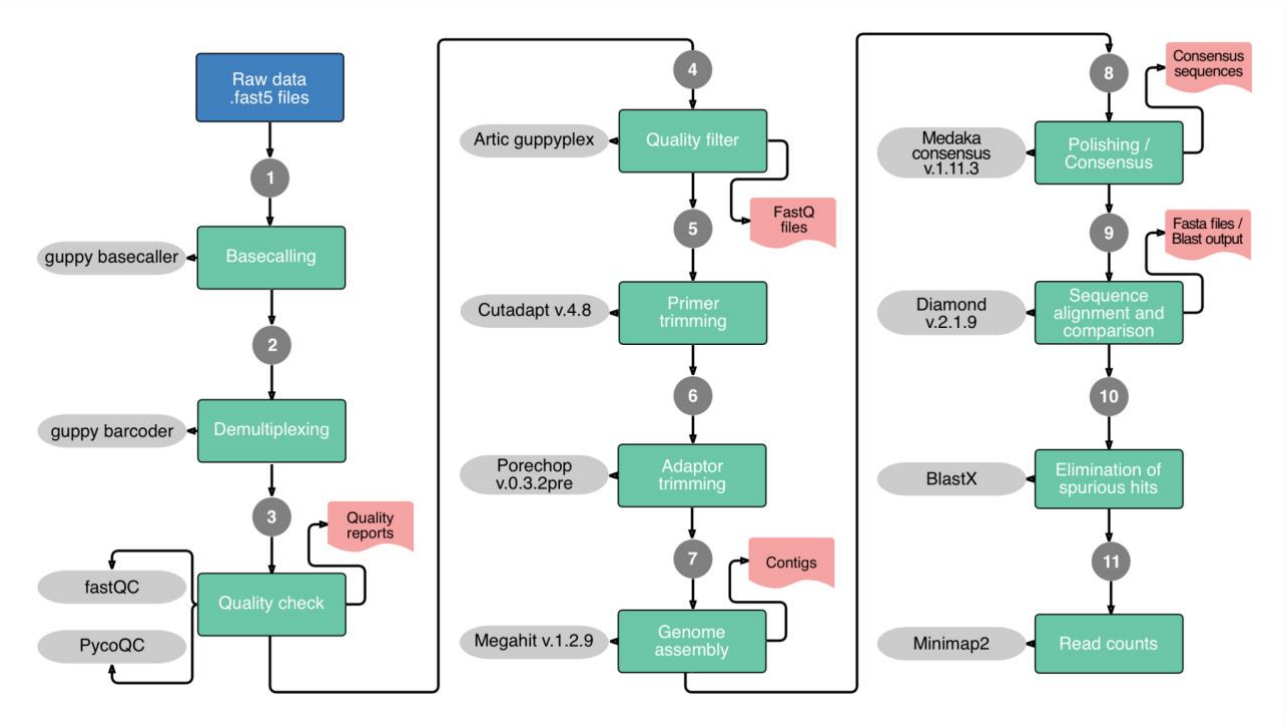
410 **Bioinformatics workflow and sequence analysis**

411 Raw sequencing reads were basecalled and demultiplexed using Guppy (Oxford Nanopore
412 Technologies, ONT). Quality control analyses were conducted with FastQC v0.12.1
413 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and pycoQC v2.5.2. Primer and
414 barcode trimming was performed using Cutadapt v4.8 and Porechop v0.3.2 (ONT). We performed
415 sensitivity analysis with three different *de novo* assemblers - Flye, Raven, and MEGAHIT - and
416 determined the latter to be the most sensitive for assembling viral sequences in our dataset (details
417 on **Supplementary Methods 2**).

418 Assemblies were polished using Medaka v1.11.3 (ONT) to generate high-quality consensus
419 sequences. These sequences were queried against the NCBI RefSeq viral protein database
420 (downloaded on May 10, 2025) using DIAMOND v2.1.9 (Buchfink et al. 2021) in blastx mode with
421 the ultra-sensitive option enabled. To minimize false positives, hits with an e-value greater than 1e-
422 10 were filtered out. For each sample, trimmed reads were remapped to viral contigs using
423 Minimap2 (Li 2018) with parameters optimized for sensitivity (-k10, -w1, -s50). Samtools (Li et al.
424 2009) and Bedtools (Quinlan & Hall 2010) were used to quantify the number of mapped reads and
425 calculate sequencing coverage, respectively. Sequences primarily assigned to viruses were further
426 analyzed via diamond blastx searches against the NCBI non-redundant (nr) database to exclude
427 potential false positives. Contigs supported by fewer than two reads, shorter than 600 nucleotides or
428 belonging to viral families present in sequencing negative controls were excluded from further
429 analysis.

430 Following previous studies in viral metagenomics (Geoghegan et al. 2021; Asplund et al.
431 2019), we applied an index-hopping filter at the genus level to remove spurious assignments.
432 Specifically, a viral contig was considered valid only if it was supported by a number of reads

433 greater than 1% of the maximum read count observed for that genus across all samples in the same
434 sequencing run. Open reading frames (ORFs) for relevant viral contigs were identified with the
435 NCBI ORFfinder web-app.
436 To further assess the zoonotic potential of each virus, we used the SpillOver web tool
437 (Grange et al. 2021). This platform employs a scoring model based on multiple virological and
438 ecological factors to provide a relative risk measure and rank viruses by their likelihood of spilling
439 over from animal hosts into human populations. Additionally, predicted protein structures of
440 putative amino acid sequences were inferred using the HHpred web application (Zimmermann et al.
441 2018). The bioinformatics pipeline is summarized in **Fig 7**.
442



443
444 **Figure 7.** Bioinformatics pipeline used in the analysis.
445

446 **Phylogenetic analyses**

447 Deduced amino acid sequences from viral contigs assigned to selected viral families were
448 subjected to phylogenetic analyses. On occasions where more than one contig was identified for a
449 given genus or species per sample, only the largest contig was included in the analysis. Reference
450 datasets were compiled by retrieving protein sequences from reference genomes for each viral
451 family (as compiled by the International Committee on Taxonomy of Viruses - ICTV),
452 supplemented with the 100 closest homologous sequences identified in the NCBI non-redundant
453 (nr) protein database. Sequence alignments were performed using MUSCLE (Edgar 2004) with
454 default parameters and subsequently trimmed with TrimAl (Capella-Gutiérrez et al. 2009) using the
455 -gappyout option. Maximum Likelihood phylogenetic trees were inferred with IQ-TREE2 v2.0.7

(Minh et al. 2020), under the best-fitting amino acid substitution model determined by ModelFinder (Kalyaanamoorthy et al. 2017). Branch support was assessed using 1,000 replicates of the Shimodaira–Hasegawa approximate likelihood ratio test (SH-aLRT) (Guindon et al. 2010).

Statistical analysis

To evaluate differences in viral community composition between intestinal and pulmonary samples, we computed pairwise dissimilarities using the Bray-Curtis index, which captures both the presence/absence and relative abundance of viral taxa. Statistical significance of compositional differences between groups was assessed using Permutational Multivariate Analysis of Variance (PERMANOVA), a non-parametric method based on permutations of the distance matrix that tests for differences in multivariate centroid locations among groups (Anderson 2008). A p-value < 0.05 was considered statistically significant. All analyses were performed in R Statistical Software v4.3.1 (R Core Team 2021) using RStudio v2024.04.2. Ecological analyses were conducted using the vegan package v2.6-8 (Oksanen et al. 2001), and data visualization was performed with ggplot2 (Wickham 2016).

Acknowledgments

This work was supported by Instituto Todos pela Saúde (ITpS), under the project A125. We thank Pan American Health Organization (PAHO) for technical-scientific and logistical support. The PAHO authors alone are responsible for the views expressed in this publication, and they do not necessarily represent the decisions or policies of the Pan American Health Organization.

Author contributions

Conceptualization: AARM, GTB, JAC; Data Curation: AFB, FRRM, JAC; Formal analysis: FRRM, GLW, JAC, PEB; Funding acquisition: AFB, BAC, VSS; Investigation: AARM, ARR, DCO, GTB, JAC; Methodology: AFB, FR, GLW, MANV, PEB, RFCS, RGS, VSS; Project administration: AFB, BAC, JAC; Resources: RFCS, RGS; Software,GLW
Supervision: AFB, JAC; Visualization: AFB, FRRM, GLW, JAC; Writing (Original Draft): AARM, GTB, JAC; Writing (Review & Editing): AARM, AFB, ARR, BAC, DCO, FR, FRRM, GLW, GTB, JAC, MANV, PEB, RFCS, RGS, VSS.

Competing interests

The authors declare no competing interests.

Data availability

The raw data were submitted to the National Center of Biotechnology Information (NCBI) under BioProject PRJNA1205703, BioSamples from SAMN46054717 to SAMN46054740; from SAMN46350146 to SAMN46350421; from SRA SRR31872231 to SRR31872254; and from SRR32069178 to SRR32069453. The sequences have been submitted to Genbank/NCBI, and can be found with the accession numbers PQ650654.2, PQ963794.2, PV975644-PV975703.

References

Almeida, M.F. et al., 2015. Fauna de morcegos (Mammalia, Chiroptera) e a ocorrência de vírus da raiva na cidade de São Paulo, Brasil. *Veterinaria e Zootecnia*, 22(1), pp.89–100.

Anderson, M.J., 2008. A new method for non-parametric multivariate analysis of variance. *Austral ecology*, 26(1), pp.32–46.

Asplund, M. et al., 2019. Contaminating viral sequences in high-throughput sequencing viromics: a linkage study of 700 sequencing libraries. *Clinical microbiology and infection: the official publication of the European Society of Clinical Microbiology and Infectious Diseases*, 25(10), pp.1277–1285.

Bazzoni, E. et al., 2024. Bat ecology and microbiome of the gut: A narrative review of associated potentials in emerging and zoonotic diseases. *Animals: an open access journal from MDPI*, 14(20). Available at: http://dx.doi.org/10.3390/ani14203043.

Bentim Góes, L.G. et al., 2022. Highly diverse arenaviruses in neotropical bats, Brazil. *Emerging infectious diseases*, 28(12), pp.2528–2533.

Branda, F. et al., 2024. Zoonotic paramyxoviruses: Evolution, ecology, and public health strategies in a changing world. *Viruses*, 16(11). Available at: http://dx.doi.org/10.3390/v16111688.

Buchfink, B., Reuter, K. & Drost, H.-G., 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature methods*, 18(4), pp.366–368.

Capella-Gutiérrez, S., Silla-Martínez, J.M. & Gabaldón, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics (Oxford, England)*, 25(15), pp.1972–1973.

Cerri, A. et al., 2023. Identification and characterization of novel alphacoronaviruses in Tadarida brasiliensis (Chiroptera, Molossidae) from Argentina: insights into recombination as a mechanism favoring bat coronavirus cross-species transmission. *Microbiology spectrum*, p.e0204723.

Claro, I.M. et al., 2021. Rapid viral metagenomics using SMART-9N amplification and nanopore sequencing. *Wellcome open research*, 6, p.241.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, 32(5), pp.1792–1797.

527 Festa, F. et al., 2023. Bat responses to climate change: a systematic review. *Biological reviews of*
528 *the Cambridge Philosophical Society*, 98(1), pp.19–33.

529 Garbino, G.S.T. et al., 2024. Updated checklist of bats (Mammalia: Chiroptera) from Brazil.
530 *Zoologia (Curitiba, Brazil)*, 41. Available at: http://dx.doi.org/10.1590/s1984-
531 4689.v41.e23073.

532 Geoghegan, J.L. et al., 2021. Virome composition in marine fish revealed by meta-transcriptomics.
533 *Virus evolution*, 7(1), p.veab005.

534 Gilbert, R. & Cliffe, S.J., 2016. Public Health Surveillance. In *Public Health Intelligence*. Cham:
535 Springer International Publishing, pp. 91–110.

536 Grange, Z.L. et al., 2021. Ranking the risk of animal-to-human spillover for newly discovered
537 viruses. *Proceedings of the National Academy of Sciences of the United States of America*,
538 118(15), p.e2002324118.

539 Guindon, S. et al., 2010. New algorithms and methods to estimate maximum-likelihood
540 phylogenies: assessing the performance of PhyML 3.0. *Systematic biology*, 59(3), pp.307–321.

541 Ikegame, S. et al., 2023. Metagenomics-enabled reverse-genetics assembly and characterization of
542 myotis bat morbillivirus. *Nature microbiology*, 8(6), pp.1108–1122.

543 Kalyaanamoorthy, S. et al., 2017. ModelFinder: fast model selection for accurate phylogenetic
544 estimates. *Nature methods*, 14(6), pp.587–589.

545 Kawasaki, J., Tomonaga, K. & Horie, M., 2023. Large-scale investigation of zoonotic viruses in the
546 era of high-throughput sequencing. *Microbiology and immunology*, 67(1), pp.1–13.

547 Letko, M. et al., 2020. Bat-borne virus diversity, spillover and emergence. *Nature reviews.*
548 *Microbiology*, 18(8), pp.461–471.

549 Li, H., 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics (Oxford,*
550 *England)*, 34(18), pp.3094–3100.

551 Li, H. et al., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford,*
552 *England)*, 25(16), pp.2078–2079.

553 Luz, H.R. et al., 2016. Ticks parasitizing bats (Mammalia: Chiroptera) in the Caatinga Biome,
554 Brazil. *Revista brasileira de parasitologia veterinaria [Brazilian journal of veterinary*
555 *parasitology]*, 25(4), pp.484–491.

556 Mackenzie, J.S. et al., 2016. The role of bats as reservoir hosts of emerging neuroviruses. In
557 *Neurotropic Viral Infections*. Cham: Springer International Publishing, pp. 403–454.

558 Miller, R.R. et al., 2013. Metagenomics for pathogen detection in public health. *Genome medicine*,
559 5(9), p.81.

560 Minh, B.Q. et al., 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference
561 in the Genomic Era. *Molecular biology and evolution*, 37(5), pp.1530–1534.

562 Minter, A., Medley, G.F. & Hollingsworth, T.D., 2024. Using passive surveillance to maintain
563 elimination as a public health problem for neglected tropical diseases: A model-based
564 exploration. *Clinical infectious diseases: an official publication of the Infectious Diseases*

Society of America, 78(Supplement_2), pp.S169–S174.

Nunes, H., Rocha, F.L. & Cordeiro-Estrela, P., 2017. Bats in urban areas of Brazil: roosts, food resources and parasites in disturbed environments. *Urban ecosystems*, 20(4), pp.953–969.

Quan, P.-L. et al., 2013. Bats are a major natural reservoir for hepaciviruses and pegiviruses. *Proceedings of the National Academy of Sciences of the United States of America*, 110(20), pp.8194–8199.

Quinlan, A.R. & Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*, 26(6), pp.841–842.

Rahimian, M. & Panahi, B., 2024. Metagenome sequence data mining for viral interaction studies: Review on progress and prospects. *Virus research*, 349(199450), p.199450.

Sasaya, T. et al., 2023. ICTV virus Taxonomy profile: Phenuiviridae 2023. *The Journal of general virology*, 104(9). Available at: http://dx.doi.org/10.1099/jgv.0.001893.

Sayler, K.A. et al., 2014. Isolation of Tacaribe virus, a Caribbean arenavirus, from host-seeking Amblyomma americanum ticks in Florida. *PloS one*, 9(12), p.e115769.

Schaeffer, R., Temeeyasen, G. & Hause, B.M., 2022. Alphacoronaviruses are common in bats in the upper Midwestern United States. *Viruses*, 14(2), p.184.

Van Brussel, K. & Holmes, E.C., 2022. Zoonotic disease and virome diversity in bats. *Current opinion in virology*, 52, pp.192–202.

Wallau, G.L. et al., 2023. The virome of bats inhabiting Brazilian biomes: Knowledge gaps and biases towards zoonotic viruses. *Microbiology spectrum*, 11(1), p.e0407722.

Wells, H.L. et al., 2022. Classification of new morbillivirus and jeilongvirus sequences from bats sampled in Brazil and Malaysia. *Archives of virology*, 167(10), pp.1977–1987.

Xu, Z. et al., 2022. Virome of bat-infesting arthropods: Highly divergent viruses in different vectors. *Journal of virology*, 96(4), p.e0146421.

Yang, X.-L. et al., 2017. Genetically Diverse Filoviruses in Rousettus and Eonycteris spp. Bats, China, 2009 and 2015. *Emerging infectious diseases*, 23(3), pp.482–486.

Zimmermann, L. et al., 2018. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *Journal of molecular biology*, 430(15), pp.2237–2243.