Ambecovirus, a novel *Betacoronavirus* subgenus circulating in neotropical bats sheds new light on bat-borne coronaviruses evolution

Gabriel da Luz Wallau[1,2,3]*, Eder Barbier[4,5], Lais Ceschini Machado[1], Alexandre Freitas da Silva[1], Yago Jose Mariz Dias[1], Alexandru Tomazatos[6], Balázs Horváth[6], Roberto D. Lins[7], Enrico Bernard[4§], Dániel Cadar[6]*

*Corresponding authors

1. Departamento de Entomologia e Núcleo de Bioinformática, Instituto Aggeu Magalhães, Fundação Oswaldo Cruz, Recife, Pernambuco, Brazil
2. Department of Arbovirology and Entomology, Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany.
3. Universidade Federal de Santa Maria (UFSM), Santa Maria, Rio Grande do Sul, Brazil.
4. Laboratório de Ciência Aplicada à Conservação da Biodiversidade, Departamento de Zoologia, Centro de Biociências, Universidade Federal de Pernambuco UFPE, Recife, Pernambuco, Brazil.
5. Departamento de Patologia, Reprodução e Saúde Única, Universidade Estadual Paulista, Jaboticabal, São Paulo, Brazil.
6. Virus Metagenomics and Evolution Group, Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany.
7. Departamento de Virologia, Instituto Aggeu Magalhães, Fundação Oswaldo Cruz, Recife, Pernambuco, Brazil - § Present address: Laboratório de Ciência em Biodiversidade, Departamento de Ecologia e Conservação, Instituto de Ciências Naturais, Universidade Federal de Lavras UFLA, Minas Gerais, Brazil

Abstract

Understanding the viral diversity harboured by wildlife is essential for effective prediction and prevention of future zoonotic outbreaks. Bats, in particular, are recognized as natural reservoirs for several zoonotic viral pathogens of high impact on public health, including coronaviruses responsible for SARS, the rabies virus, Marburg, Ebola, Nipah and Hendra viruses. However, the large extent of bat viruses remains unexplored, especially in highly biodiverse regions of the Neotropics such as Brazil. We used a meta-transcriptomic to characterize new virus genomes found in blood, oral and anal samples collected from cave- and non-cave bats from Northeast Brazil. From a total of 19 coronavirus-positive bats, we have assembled two complete genomes of a new *Betacoronavirus* subgenus, named *Ambecovirus* (American betacoronavirus). The subgenus herein described is phylogenetically placed between the *Sarbeco-/Hibeco-/Nebeco*virus and the *Merbeco-/Embecovirus* clades, being basal to the former. While the conserved S2 region of the spike protein retained hallmark domains, including HR1 and HR2, the S1/S2 cleavage site and the furin cleavage site, the S1 region consistently displayed only the N-terminal domain. The receptor-binding domain could not be identified due to high dissimilarity relative to known congeneres. The detection of *Ambercovirus* in sympatric *Pteronotus gymnonotus* and *Carollia perspicillata* bats suggests interspecies transmission. Longitudinal sampling confirmed persistent *Ambecovirus* infection in *P. gymnonotus* over multiple years and virus dispersion at a minimum distance of 270 km between caves. The present study confirms that viral diversity in neotropical hosts remains largely unknown not just in Brazil but, likely, in the other countries of the region, supporting the need for a systematic approach to virome exploration and analysis followed by *in vitro* experimentation to assess zoonotic potential.

Keywords: meta-transcriptomics, *Coronaviridae*, virus discovery, wildlife, bat, zoonotic potential, comparative genomics

Introduction

Viruses are obligate parasites and represent some of the most widespread and biologically diverse entities around the planet (1,2). Due to this parasitic lifestyle, many viruses cause diseases in their hosts, but as our knowledge advances on the virosphere (viral communities), evidence points to a consensus that the majority of viruses are non-pathogenic with a small fraction with clear pathogenic profile (3–5). Still, such evidence is based on a very limited - but expanding - view about the known virosphere (3,6). Regarding pathogenic viruses, a growing body of evidence suggests that several natural host reservoirs are tolerant to infections and experience little or no disease likely due to the long-term host-virus coevolution, and viral tolerance profiles evolved along millions of years of evolution (7). Bats (Order Chiroptera), the only mammals capable of sustained flight, are special examples of reservoirs of a number of known pathogenic viruses and, yet, appear mostly lightly or even unaffected by them (8). This group of mammals have gone through multiple adaptations to perform active flight, including evolving differential immunological responses, which are linked to their virus tolerance profile (9,10).

Bats are the second largest order of mammals (~22% of known species) and have a worldwide distribution (11). Aside from their ecological roles as keystone-species (e.g. in plant pollination and seed dispersal), they are recognized as natural reservoir hosts of several viruses of high zoonotic importance that belong to different viral families such as *Coronaviridae*, *Paramyxoviridae*, *Rhabdoviridae* and *Filoviridae* (12). Our current understanding of their role in maintaining the circulation of several zoonotic virus species relies mainly on longitudinal surveillance efforts focused on rabies, Hendra and Nipah viruses in the Americas, Europe, Asia and Oceania (13,14). More recently, it also became evident that bats are the natural reservoirs of coronaviruses, including close relatives of human pathogens like SARS-CoV, SARS-CoV-2 and MERS-CoV betacoronaviruses, as well as of other *alphacoronaviruses*, such as HCoV-229E and HCoV-NL63 associated with seasonal, mild respiratory disease in humans (15–18). Most of the evidence on bats as natural reservoirs of these viruses comes primarily from low-throughput molecular methods that can capture a limited amount of the viral genetic diversity (14,19). More recently, unbiased high-throughput sequencing (metagenomics and meta-transcriptomics) has reshaped our understanding of viromes hosted by bats and by other host species, allowing hypothesis-free explorations without *a priori* knowledge of existing genetic diversity (5,20). These studies have corroborated many findings about main reservoir groups, but also uncovered new host groups and a hidden diversity of viral lineages (6).

Brazil is ranked first among the 17 megadiverse countries (21), covering six terrestrial biomes and the highest number of endemic species. With yet-uncharacterized viromes and microbiomes, plus an increasing anthropogenic pressure represented by deforestation and human encroachment in once pristine ecosystems, Brazil has been ranked as one of the most important potential hotspots for the emergence of new viral pathogens (22–25). With 186 known species in nine families, Brazil accounts for 12.5% of the world's bat fauna (1.487 species) (26). Despite such bat species richness and diversity,

nearly half of all bat species in Brazil have not been yet investigated as virus hosts (19). Moreover, most of the virus detection research performed so far in the country relied mainly on low throughput and targeted methodologies. Thus, the results obtained so far have been biased by preferential sampling of known reservoirs of rabies virus (19). Additionally, few longitudinal studies were performed using varied sampling strategies across different biomes in Brazil. Coronaviruses, for example, have been detected in neotropical bats from Mexico, Brazil, Argentina, Peru, Trinidad and Tobago and other countries, primarily through PCR and Sanger sequencing (27–37,37,38). These findings revealed a predominance of alpha- over betacoronaviruses infecting neotropical bats (39,40). So far, only one genomic study has been performed in Brazil reporting two betacoronaviruses, the first a MERS-CoV-related virus, while the second showed no clear phylogenetic clustering with known *Orthocoronavirinae* genera (41). These findings underscore the substantial gaps in our understanding of the bat virome and its zoonotic potential in Brazilian ecosystems as well as on coronavirus evolution.

Here we employed meta-transcriptomic untargeted high throughput sequencing for virus discovery in cave- and non-cave bats from Northeastern Brazil, particularly from the Caatinga drylands and from the Atlantic Forest biomes. We recovered complete *Betacoronavirus* genomes and compared their similarity to known coronaviruses using a comparative analysis to characterise the architecture of genomes, spike protein domains and revisit the deep evolutionary roots of the *Betacoronavirus* genus.

Material and Methods

Bat sampling

Sampling was performed from May 2019 until January 2023 across 12 sites in Northeastern Brazil (**Figure 1**). The initial collection, associated with a natural history and ecology survey, involved whole blood samples (30–50 µL) mixed with 100 µL of RNAlatter (Thermo Fisher Scientific). From 2020 onwards, oral and anal swabs were collected and directly enclosed in Eppendorf tubes and placed in liquid nitrogen containers without any buffer solution. Swabs were then transferred to -80℃ freezers after nitrogen liquid containers reached stationary laboratories and remained there until sample processing.

Bats were captured using harp traps placed at cave entrances at 5:30 p.m. for one hour. In urban areas, bats were captured using four mist nets (12 m x 2.5 m each) set for a 4-hour period starting at dusk. Trapped/netted individuals were carefully removed, weighted and identified at the species level following specific literature (42–45). The bat's age was estimated based on the metacarpal epiphyseal cartilages (46), with individuals classified as adults (closed epiphysis) or subadults (open epiphysis). The reproductive stage of bats was assessed based on the presence of secondary sexual characteristics. Female individuals were classified into three categories: pregnant, lactating, or inactive (non-reproductive). Males were considered reproductively active when they exhibited visibly enlarged testes. All bat handling was conducted by trained personnel with protection equipment, including NFP2

masks, face shield, gloves and lab coat. Fieldwork permit was issued by the Instituto Chico Mendes de Conservação da Biodiversidade–ICMBio through SISBIO authorizations (68992-1, 68992-2, 68992-3, 77600-1, 83959-1, and 83959-2) and approval by the Ethics Committee on Animal Care (CEUA) of the Universidade Federal de Pernambuco (Process numbers 114/2019 and 092/2021). Moreover, the study was also registered in Brazil's Sistema Nacional de Gestão do Patrimônio Genético e do Conhecimento Tradicional Associado (SisGen) under the number A35D254.

Sampling processing and metagenomic sequencing

We applied a virus-optimized metagenomic sequencing protocol adapted from established approaches for unbiased viral detection (46) to characterize the virome of the cave- and non-cave bats. Specimens were pooled by species, location, and collection date (1–4 individuals per pool). To reduce host and microbial nucleic acid background, samples underwent a viral particle enrichment protocol. Blood samples were diluted in sterile PBS, and swab samples were eluted in PBS, then incubated with Proteinase K at 50 °C for 30 minutes. The supernatant was subsequently filtered through a sterile 0.45 µm membrane (Millipore) to eliminate cellular debris and larger contaminants. Filtered aliquots (250–400 µL) were treated with a cocktail of nucleases, including Turbo DNase (Ambion), Baseline-ZERO DNase (Epicenter), Benzonase (Novagen), and RNAse One (Promega), to selectively degrade unprotected DNA and RNA, enriching for viral particles. Total nucleic acids were extracted using the MagMAX Viral RNA Isolation Kit (Life Technologies) for blood and the QIAamp Viral RNA Mini Kit (Qiagen) for swab samples, according to manufacturer protocols. Each sequencing batch included a negative water control to monitor for reagent contamination. Reverse transcription and cDNA synthesis were performed using an octamer random primer and the amplified DNA was used for library preparation with the QIAseq FX DNA Library Kit (Qiagen), incorporating dual-index barcoding to enable multiplex sequencing. Library concentrations were assessed using Qubit fluorometry and the Agilent Bioanalyzer. Sequencing was performed on the Illumina NextSeq 2000 platform (2×100 cycles).

Bioinformatic analysis

Assembly and viral contigs binning

Raw sequences were filtered for quality (Phred score ≥20) and length (minimum 50 bases) and trimmed using Trimmomatic 0.39 to remove polyclonal and low-quality reads (47). To reduce artefactual read duplication introduced during random RT-PCR, deduplication was performed using the 'dedupe.sh' tool from the BBTools suite (v39.01) (48). Ribosomal RNA was removed by mapping clean, deduplicated reads to the SILVA database v138.1 (49). High quality, filtered sequences were then *de novo* assembled with Megahit v1.2.9 (50) with a minimum length limit of 400 bases. Contigs were compared against the NCBI non-redundant (nr) database (2024.03.16.) using DIAMOND v2.1.9 blastx with E-value cutoff of 0.001 for high sensitivity and to reduce false positives (51). To retrieve standardized assembly genomic

metrics and consensus genomes, a complementary reference-based assembly approach was employed. This involved passing the newly complete and annotated genomes recovered from the *de novo* assembly strategy and the raw sequencing reads to ViralFlow v1.0 (52). CoverM tool (https://github.com/wwood/CoverM) was also used to obtain genomic metrics (53).

Contigs extension and annotation

The contigs were visually validated using the Geneious Prime 2024 mapping algorithm and extended through several iterations where possible. ORFs were predicted with Prodigal v2.6.3 (54) and MEGAN6 (55), then manually corrected as needed. Special attention was given to potential virus-specific coding strategies (e.g., ribosomal shifting, overlapping reading frames, transcriptional slippage, leaky scanning, alternative splicing) by using the closest available reference sequence in CLC Genomics Workbench 24 (Qiagen). All genomic sequences generated during this study have been submitted to GenBank and accession numbers will be updated here as soon as they become available. The associated raw sequencing datasets are publicly accessible via the Sequence Read Archive (SRA), linked to Bioproject ID PRJNA1291827.

Homologous sequences recovery

To investigate the evolutionary relationships of the identified coronaviruses, we performed homologous sequence searches using BLASTn and BLASTp at NCBI (last accessed June 2025). To place these new coronaviruses within the *Orthocoronavirinae* phylogenetic tree we used the amino acid sequences of the hallmark genes such as RNA dependent RNA polymerase (RdRp) for homologous searches. This included the amino acid sequence of reference viral genomes from the *Orthocoronavirinae* subfamily classified by the International Committee on Taxonomy of Viruses (https://ictv.global/). Phylogenetic reconstruction was based on RdRp sequences, since it reflects vertical ancestry and is less prone to recombination than other genomic regions (56). Additionally, to explore conserved domains in rapidly evolving proteins, we conducted amino acid homology searches of spike glycoproteins, which mediate host cell receptor binding. On the other hand, to study more recent evolutionary relationships we also performed blastn searches at the NCBI Virus (https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/) and the ZOVER databases (https://www.mgc.ac.cn/cgi-bin/ZOVER/main.cgi).

Multiple sequence alignment and phylogenetic reconstruction

Recovered and generated sequences were aligned using MAFFT v7 (57) and inspected using Aliview (58) both at the amino acid and nucleotide level. Phylogenetic reconstruction was performed

using IQTREE v2 (59) after model selection with ModelFinder (60) implemented in IQTREE. Branch support was accessed using aLRT and ultrafast bootstrap (UFboot) with 1000 replicates. Beast v1.10 (61) was used to perform a Bayesian phylogenetic reconstruction for the *Orthocoronavirinae* phylogenetic tree to reassess specific branch support.

In silico protein characterization

In order to obtain more detailed insights about the betacoronavirus genome recovered we sought to characterize the conserved domains of the spike protein through searches at the Conserved Domain Database (CDD search - https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). To further explore and attempt to align the spike protein of the viruses sequenced in this study with other coronavirus spikes we also used the COBALT software for constraint-based alignment tool (62).

Three-dimensional models for the trimeric structure of coronavirus spike protein were generated using AlphaFold2 (63) and the SwissModel server (64). Multiple sequence alignment in AlphaFold2 was performed using the mmseqs2_uniref_env and unpaired_paired parameters and the alphafold_multimer_v3 model. The remaining parameters were set to default.

Electrostatic potentials for the NTD regions of HCoV-HKU1 and other coronavirus models were calculated using the linearized Poisson-Boltzmann equation using the APBS software (65). PDB2PQR (66) was used to generate the PQR formatted files from the PDB coordinates by assigning partial atomic charges for the proteins using the PARSE forcefield at pH 7.0 and the solvent described as a dielectric constant of 78 and saline concentration of 0.15 M. The low dielectric cavity was set to 4.

Results

Seventeen bat species were sampled at 12 sites (11 caves and one urban area), totalling 452 individuals and 712 biological samples processed (**Figure 1** and **Supplementary Table 1**).

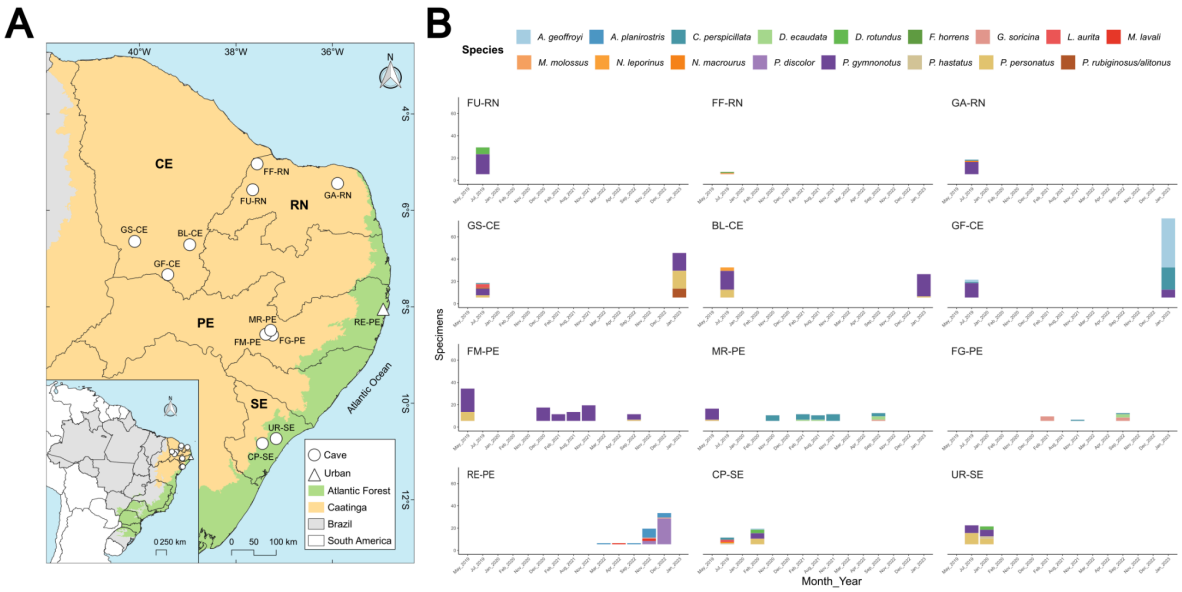


Figure 1 - Bat sampling performed during the study period. **A** - Sampling sites including 11 caves (circles) and 1 urban (triangle) site in the Northeast part of Brazil. Colors denote two of the six Brazilian biomes; orange (Caatinga) and green (Atlantic forest). **B** - Number of specimens collected per sampling site during this study. The first two letters of the site acronym stands for the name of the cave (GS -Gruta do Sobradinho, BL - Boqueirão de Lavras II, GF - Gruta do Farias, FF - Furna Feia, FU - Furna do Urubu, GA - Gruta do Arnoud, MR - Meu Rei, FM - Furna do Morcego, FG - Furna do Gato, UR - Urubu, CP - Casa de Pedra) while the following two letters for stand for the state where the cave was located (CE - Ceará, RN = Rio Grande do Norte, PE - Pernambuco and SE - Sergipe).

We obtained two complete coronavirus genomes for samples from *Pteronotus gymnonotus* (Mormoopidae): Pgymn_N107_15 (30.414 bp) and Pgymn_N107_12 (30.423 bp) with average coverage depth of 621x and 314x (**Figure 2 and Supplementary Figure 1 and Table 2**). In addition, other seven partial genomes with a variable coverage breadth varying from 4 to 76% were obtained from the species *P. gymnonotus* and *Carollia perspicillata* (Phyllostomidae) (**Table 1**, **Supplementary Figure 2**). Full genomic annotation revealed a consistent gene order and gene length of the *Betacoronavirus* genus (**Figure 2A, Supplementary Figure 3**). The predicted spike protein length ranged from 1449 to 1453 amino acids and showed a more variable S1 and conserved S2 subunits (**Figure 2B**). Conserved domain analysis identified the N-terminal domain (NTD) within the S1 region including the trimer interface, but no receptor-binding domain (RBD) signature was detected (**Figure 2B**). Notably, despite the presence of the NTD, no clear amino acid alignment of the S1 region could be established when compared to homologs retrieved via BLASTp against the NCBI nr database, including well-characterized reference spike proteins from SARS-CoV, SARS-CoV-2 and MERS-CoV) (**Figure 2C** and **Supplementary File 1**). In contrast, conserved domains within the S2 region were identified, including the S1/S2 cleavage region, S2 cleavage site with the internal fusion peptide and the Heptad Repeat domains HR1 and HR2 (**Figure 2B and D**). Interestingly, the two complete coronavirus genomes were nearly identical except by the spike protein. Considering the 1449-1543 predicted amino acids, the

pairwise protein similarity is ~79%, meaning that 1147 amino acid residues are identical between the two spike proteins. The majority of distinct spike amino acids are at the more variable S1 region that encompass the NTD and CTD regions (**Supplementary File 2**).

Table 1 - Descriptive statistics of complete and draft coronavirus genomes recovered in this study and metadata associated with the samples.

Pool name	Species	Nº specimens	Contig length	Total Pass Filter Reads	Mapped Read Count	Mean Coverage Depth	Sex	Age	Reproductive stage*
N107_10	*P. gymnonotus*	1	5,334	2908050	5131	17.54	Male	Adult	Active
N107_11	*P. gymnonotus*	1	10,882	3206766	13073	45.04	Male	Adult	Inactive
N107_12	*P. gymnonotus*	1	30,422	3400908	101845	336.54	Male	Adult	Inactive
N107_13	*P. gymnonotus*	1	8,477	1178662	1865	6.25	Male	Adult	Inactive
N107_15	*P. gymnonotus*	1	30,414	3294633	182751	631.17	Male	Adult	Inactive
N107_17	*P. gymnonotus*	1	4,139	3647063	1927	6.65	Male	Adult	Inactive
N107_31	*P. gymnonotus*	1	23,261	3977163	37408	127.28	Male	Adult	Active
N107_32	*P. gymnonotus*	1	12,301	3789009	21238	64.27	Male	Adult	Active
N107_34	*C. perspicillata*	1	3,682	4793598	2147	7.32	Female	Adult	Inactive
N152_004	*P. gymnonotus*	2	9,276	2428400	858	3.54	Female/Male	Adult/Adult	Inactive/Active
N152_005	*P. gymnonotus*	2	11,358	3199622	1909	7.14	Male	Adult	Active
N152_006	*P. gymnonotus*	2	2,032	3287243	306	1.29	Male	Adult	Active
N152_007	*P. gymnonotus*	2	2,049	3496616	135	0.50	Male	Adult	Active
N152_013	*P. gymnonotus*	2	11,366	3140953	2977	11.06	Female/Male	Adult/Adult	Inactive/Inactive
N152_014	*P. gymnonotus*	2	15,974	755285	1262	5.24	Male/Male	Adult/Adult	Inactive/Active
N152_016	*P. gymnonotus*	2	14,587	3216751	8191	30.13	Male/Male	Adult/Subadult	Active/Inactive
N152_018	*P. gymnonotus*	2	2,112	1802108	272	0.90	Male/Male	Adult/Adult	Active/Inactive
N152_080	*P. gymnonotus*	1	1,207	1002122	65	0.20	Male	Adult	Active
N152_160	*P. gymnonotus*	1	1,705	1152208	103	0.38	Female	Adult	Lactating

* Female individuals were classified into three categories: pregnant, lactating, or inactive (non-reproductive). Males were considered reproductively active when they exhibited visibly enlarged testes.

Structural modeling of the spike protein proved to be challenging, especially due to low sequence identity at the S1 region. None of the methods used reached acceptable quality metrics. AlphaFold3 yielded a pTM score of only 0.3, while SwissModel template-based strategy performed marginally better, achieving a GQME of 0.45 using the PDB ID 9BSW (67) of HCoV-HKU1 (**Figure 3A**). Nevertheless, a common feature among the models was the prediction of an NTD β-sheet motif. SwissModel template matching showed nearly 30% identity to lineage A betacoronavirus HCoV-HKU1

and HCov-OC43 for Pgymn_S15 and Pgymn_S12, respectively, two members of the *Embecovirus* subgenus within the *Betacoronavirus* genus. Both viruses use a β-sheet structured shallow groove in the NTD to bind to sialic acids containing-receptors. These residues are located within the first 210 residues of the S1 subunit for the HCoV-HKU1 and HCoV-OC43. A three-dimensional superposition of the predicted NTD structure of Pgymn_S15 by SwissModel and PDB ID 9BSW of HCoV-HKU1 is shown in **Figure 3B**. The β-sheet motif, which is also predicted for Pgymn_S12, is three-dimensionally well aligned suggesting ambecov S1-NTD would be contained within the same residue range. Prediction of the CTD was contrasting and poorly structured in all cases, as shown for Pgymn_S15 in **Figure 3C**. Although the CTD core secondary structure is predicted, there is a massive three-dimensional misalignment outside of the CTD core region. Based on this rather poor structural comparison to HCoV-HKU1 and HCov-OC43, ambecov's CTD would likely lie around residues 310 to 600 (**Figure 3C)**.

Another interesting fact about the spike protein from ambecov is that its total net charge at pH 7 is predicted to be equal to -30 *e*, the same charge adopted by HCoV-HKU1's spike protein. It is worth noting that it dramatically contrasts to the charge of spike proteins from SARS-CoV-2, which tends to be around -6 *e*. Electrostatic properties of coronaviruses proteins have been well documented to play a major role in receptor, furin and RNA recognition, as well as immune evasion (68–71). Given the net charge similarity of ambecov and HCoV-HKU1 and its reasonably well-aligned NTD region, we have calculated the electrostatic surface potential for this region in both molecules. **Figure 3D** shows a similar pattern in the sialic acid β-sheet groove for the HCoV-HKU1 and the corresponding region in the model, supporting a similar functional role.

Phylogenetic reconstruction based on conserved *Nidovirales* proteins domains (3CLpro, NiRAN, RdRP, ZBD and HEL1 domains), including representatives of the four recognized genera (Beta, Alpha, Gamma and Deltacoronaviruses) of the Orthocoronaviridae subfamily placed the newly discovered genomes within the *Betacoronavirus* genus (**Figure 4A**). Moreover, these genomes clustered as a new, distinct clade, basal to the Sarbeco, Hibeco and Nobecovirus subgenera, having relatively high node support of 71.7 (aLRT), 83% (UFboot) and 0.99 (posterior probability) (**Figure 4A**). Together with a previously reported draft genome (BetaCoV_UNIFESP_unmBSS) the new genomes clustered with high node support (aLRT and UFboot = 100%) forming a novel subgenus we propose as *Ambecovirus* (American Betacoronavirus) (**Figure 4B**). Amino acid similarity thresholds within this subgenus at the 3CLpro, NiRAN, RdRP, ZBD and HEL1 domains ranged from 80 to 100%, while intergeneric similarity varied from 69% to 70% (**Supplementary Table 3**). Moreover, a more fine grained analysis of the nucleotide sequence revealed two main clades: Clade I (branch support = 99.3 aLRT and 100% UFboot) included all genomes generated in this study, a partial sequence obtained from *Pteronotus davyi* (Mexico, 2012), seven partial sequences from *Pteronotus parnellii* (Costa Rica, 2011-2016) and a basal sequence obtained from glossophagine bats (KX284064, Brazil, 2010) (**Figure 4B**); Clade II (branch support = 99.3 aLRT and 100% UFboot) comprised of partial sequences *Artibeus* spp. and *C. perspicillata* along with a draft genome (BetaCoV_UNIFESP_unmBSS) from *Artibeus lituratus* sampled in Ceará, Brazil, in 2023

(**Figure 4B**). This clade distinction was further supported by similarity analysis of the RdRp partial (346 bp). Clade I and II showed 69-100% and 96-100% intraclade nucleotide similarity respectively, while interclade similarity ranged between 69% and 75% (**Supplementary Table 4**).

The sequences generated in this study originate from *P. gymnonotus* sampled multiple times in the same cave (Furna do Morcego, in Pernambuco state) from December 2020 to November 2021, but also from two other caves (Gruta do Farias, and Boqueirão das Lavras II) both located in Ceará state, where the same species was found infected. These two caves are located ~270 km distant from the caves in Pernambuco. Moreover, the *Ambecovirus* sequence was detected in November 2020, in *C. perspicillata* at the Meu Rei cave, also in Pernambuco, and less than 15 km from Furna do Morcego cave (**Figure 5**).

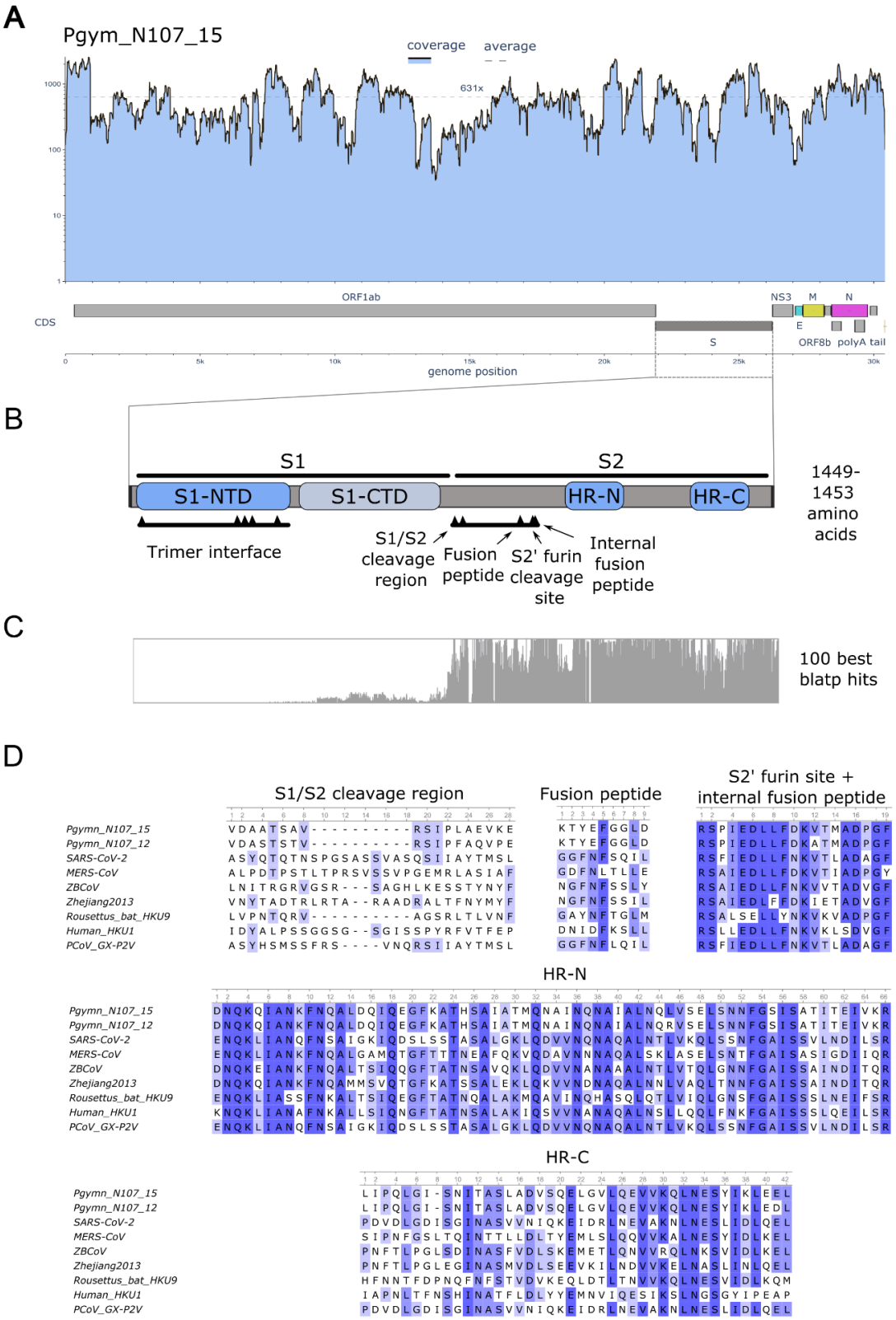


Figure 2 - Genomic features and spike protein domain annotation of the *Ambecovirus* full genome recovered in this study. **A** - Coverage breadth, depth and complete genomic annotation of the new betacoronavirus genome; **B** - Schematic of the spike protein highlighting the S1 and S2 subunits, including conserved domains; **C** - Amino acid alignment of the *Ambecovirus* spike

protein against the top 100 BLASTp hits, illustrating divergence in the S1 region; **D** - Amino acid alignment of the S2 domains found in the CDD search including the spike protein of the two complete genomes obtained in this study and reference coronavirus genomes.

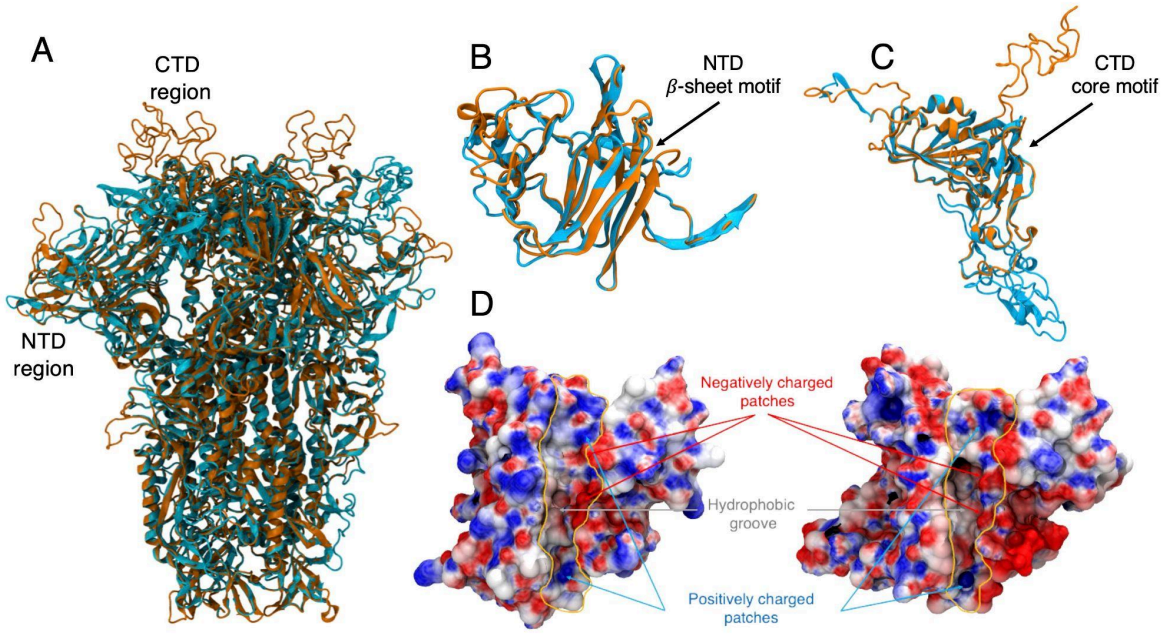


Figure 3 - Superposition of a SwissModel predicted structure (orange) of the spike protein from ambecov based on the PDB ID 9BSW template from HCoV-HKU1 (cyan) depicted in cartoon model, and comparison of the electrostatic surface potential for the NTD region. **A** shows superposition for the spike protein in its trimeric form. **B** shows the superposition for the NTDs highlighting the predicted β-sheet motif for the model (orange). **C** shows the superposition for the predicted CTD region. The putative CTD region for the model (orange) is shown as highly unstructured, the result of a massive three-dimensional misalignment outside of the core region due to a low sequence identity between template and model. **D** shows similarity in the electrostatic surface potential for sialic acid binding β-sheet groove in the NTD from HCoV-HKU1 (left) and the corresponding region in the ambecov model (right) encompassed within thin orange lines. (Electrostatic potential ranges from -5kT/*e* (red) to +5 kT/*e* (blue) and it is plotted onto the van der Waals surface).

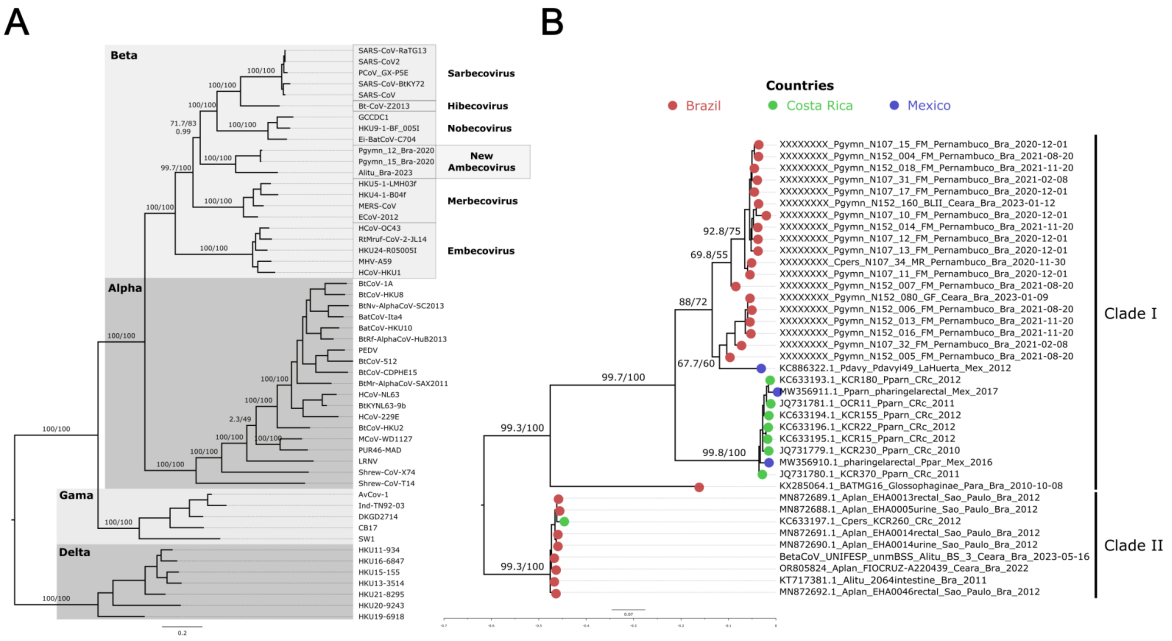


Figure 4 - Phylogenetic reconstruction of the two complete coronavirus genomes characterized in this study and representative reference coronavirus genomes at the genus level. **A** - Maximum likelihood phylogenetic reconstruction based on amino acid sequences of *Nidovirales* conserved domains (3CLpro, NiRAN, RdRP, ZBD and HEL1 domains) - **Supplementary File 3**. **B** - Total evidence nucleotide tree based on the RdRp coding region including partial (439 bp) and full RdRp nucleotide coding region (~2790 bp) of the coronaviruses sequenced in this study - **Supplementary File 4**. Branch support values above nodes were calculated by aLRT/ultrafast bootstrap and posterior probability. Tip colors denote the country of sampling.

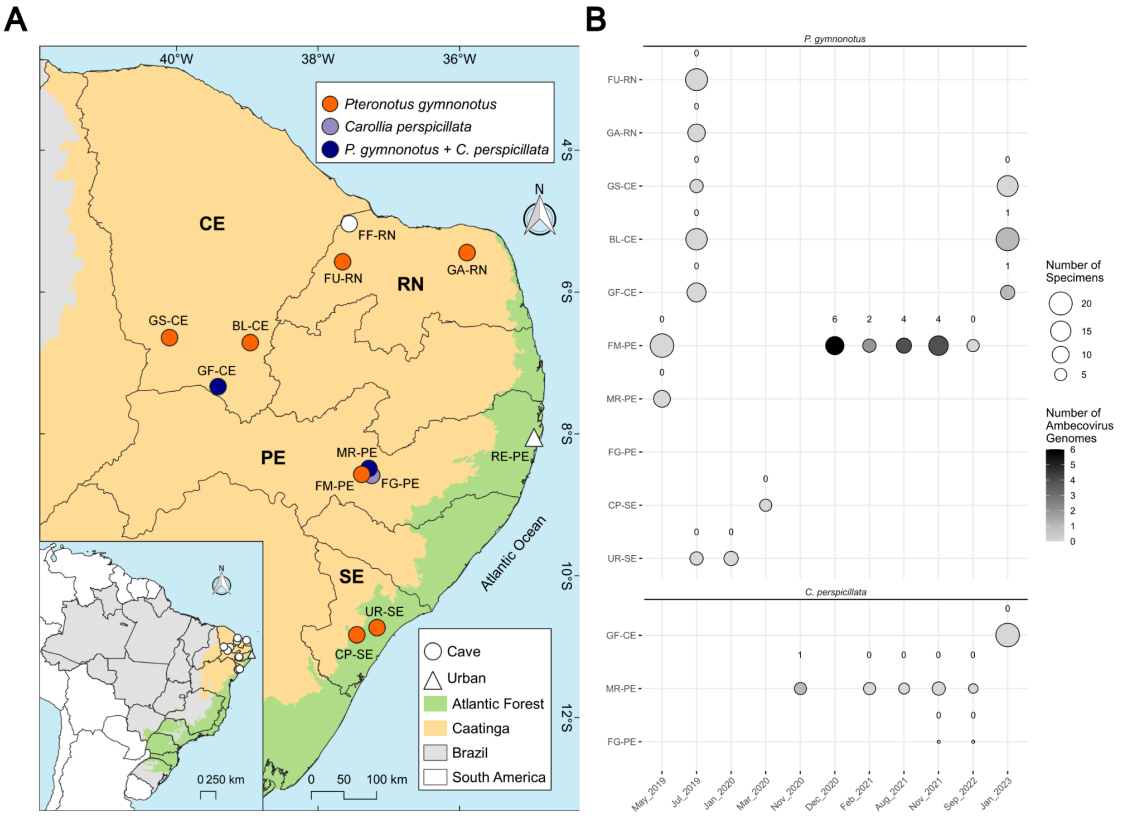


Figure 5 - Spatio-temporal distribution of ambecovirus-positive samples obtained from *P. gymnonotus* and *C. perspicillata*. **A** - Map of the caves and urban sampling sites visited along the study showing colored caves where *P. gymnonotus, C. perspicillata* or

both species co-occur. **B** - Number of individual samples per cave during the study period, including the number of partial or complete genomes recovered from metagenomic data.

Discussion

Bats are known hosts of important zoonotic pathogens including coronaviruses, rhabdoviruses (rabies), paramyxoviruses (Nipah and Hendra) and filoviruses (Ebola and Marburg) (3). Despite that, efforts to characterize bat viromes are heterogeneous with highly biodiverse areas still understudied. Among these areas, the megadiverse biomes of the Neotropics, and particularly of South America, stands out: there, almost half of all known bat species were never screened for the presence of viruses. Moreover, the importance and impact of the rabies virus on the regional herds and on human health contributed to a clear bias in favour of its known host reservoirs (19). Additionally, the overwhelming majority of investigated bat species were screened using low throughput molecular techniques that rely on *a priori* knowledge and hence are able to detect a low viral diversity (3,19). To address these gaps, we conducted viral particle enrichment and meta-transcriptomic sequencing of cave and non-cave bats from sites in two Brazilian biomes (the Atlantic forest and the Caatinga drylands). This approach led us to the discovery of a novel *Betacoronavirus* subgenus, named *Ambecovirus*, circulating in multiple bat species across Brazil and the Americas.

The majority of zoonotic coronavirus lineages belong to *Beta-* and *Alphacoronavirus* (16). Betacoronaviruses are grouped into five known subgenera, from early diverging *Embecovirus* to more recent diverging ones namely, *Merbecovirus*, *Nobecovirus*, *Hibecovirus* and *Sabercovirus*. Embecoviruses have rodents as main reservoirs, while bats are the main reservoirs of the other four known subgenera (17,72). *Merbeco* (Vespertilionidae-specific) and *sarbecoviruses* (Rhinolophidae-specific) are known to spillover to other species, driving local outbreaks to large epidemics (72) suggesting viruses from these subgenuses are more prone to cross-species transmission, while *nobeco-* and hibecoviruses have been so far only detected in bats of the families Pteropodidae and Hipposideridae, respectively (73). Our phylogenetic reconstruction based on complete RdRp sequences revealed a monophyletic clade branching as a sister group of *Nobeco-*, *Hibeco-* and *Sarbecovirus* subgenera. Although partial Ambecovirus-like sequences have previously been reported in neotropical bats, their placement within the *Betacoronavirus* genus remained uncertain (18,33,40,41). Only one recent study (2025) recovered a draft ambecovirus-like genome so far (41). The position of ambecoviruses within the *Betacoronavirus* genus, and their detection being limited so far to neotropical bats (Mormoopidae and Phyllostomidae), implies long-term host-virus coevolution and regional isolation, consistent with broader studies that pointed out that Merbecoviruses and Alphacoronaviruses in the Americas are phylogenetically distinct and less prone to cross-species transmission than from their Old World counterparts (18). More broadly, the current data suggests that an ancient split occurred in the *Betacoronavirus* genus separating ambeco- from nobeco-, hibeco- and sarbecoviruses in the Americas

while these last three shared an Old World common bat ancestor. Interestingly, this finding further corroborates other broad scale studies on coronavirus evolutionary history which showed that merbecoviruses and alphacoronaviruses infecting bats in the Americas are isolated from the African and Asian coronaviruses (18). Our data supports regional variation in the long-term evolution, but at a more recent time we detected one potential host-switching event between distantly related species (from two different neotropical bat families - Mormoopidae and Phyllostomidae) or this detection in multiple bats species suggests that ambecoviruses have broad host range infecting neotropical bats. Further studies are warranted to evaluate these hypotheses.

Evolution and zoonotic potential

The detection of a new *Betacoronavirus* subgenus raises a number of questions about its evolution and zoonotic potential. The full genomic sequence recovered in this study allows us to perform an in depth genomic annotation and characterization of the predicted proteins to test some hypotheses. Betacoronaviruses differ in the composition and order of genes at the 3' terminal region, after the spike protein. The envelope, matrix and nucleoprotein are located downstream to each other followed by hypothetical proteins in *Nobecovirus* and *Merbecovirus,* while envelope and matrix are in tandem and separated from the nucleoprotein by two to five hypothetical proteins in *Sarbecovirus* and *Hibecovirus* (https://ictv.global/report/chapter/coronaviridae/coronaviridae/betacoronavirus). The *Ambecovirus* subgenus revealed a synteny similar to *Nobecovirus* suggesting conservation of the gene order of the ancestral lineage (**Supplementary Figure 3**). Using available protein conserved domains and structural information for a number of medically important Betacoronaviruses, we characterized in more detail the spike protein responsible for receptor binding and a key molecule responsible for the infection capacity of coronaviruses. The spike protein is divided into two subunits, the more variable S1 region and the more conserved S2 one. S1 is further divided into S1-NTD and S1-CTD, while the second is mainly responsible for the fusion of the viral particle with the host cell membrane in merbeco and sarbecoviruses, the first can also bear the receptor binding bind domain to sialic acid receptors in embecovirus (74). The S1-CTD contains the receptor-binding domain and mediates receptor interaction in Merbecovirus and Sarbecovirus, but it also plays an ancillary role in Embecovirus by stabilizing the spike trimer and modulating the conformational transition between pre-fusion and post-fusion states (74). These two regions are also the main targets of neutralizing antibodies and because of the strong diversifying selection pressure, they normally evolve faster than the S2 region (75). Interestingly, *Ambecovirus* spike proteins characterized in this study showed a NTD identifiable conserved domain but lacked an identifiable RBD domain and RBM motif. Moreover, there is a complete lack of alignment of the spike S1 region with any other spike proteins available in the databases, including those of the well characterized human pathogens (MERS-CoV and SARS-CoV-2). Such pattern was already observed between the S1 subunit of different Orthocoronaviridae genera that share little sequence similarity, yet it departs from the

generally accepted view that congeners share higher sequence similarity as reported by Li (75). Therefore, the NTD and RDB region of the *Ambecovirus* are highly divergent at the amino acid level.

In order to further investigate if structural homology still exists between the S1 NTD and RBD region among Ambecov and the other Betacoronavirus subgenuses we sought to perform 3D modelling of the entire spike protein recovered. While the S1-NTD 3D modeling is compatible with a β-sheet shallow groove also found in other two Embecovirus spike 3D structures, the S1-CTD showed distinct conformations at every modelling attempt. In addition, ambecov's CTD does not contain sequence equivalence to the conserved residues found in ACE2 binding sarbecoviruses, such as Y449, G496, N/Y501 and Y505. Unlike sarbecoviruses, the CTD in HCoV-HKU1 and HCoV-OC43 plays ancillary roles, such as structural stabilization of the trimeric structure, modulation of pre-fusion to post-fusion conformational transitions, and immune evasion through glycan shielding (16,74). Hence, it is possible that the spike protein interaction with the receptor in these bats either occurs through different contact regions and/or uses sialic acid cell receptors. Another important implication is that, due to high sequence divergence, the ambecoviruses may not be able to infect human cells. However, both hypotheses must be rigorously tested using *in vitro* and *in vivo* assays to precisely evaluate its zoonotic potential. Once ambecoviruses infects both sylvatic and synanthropic bats and cohabitation with other hosts including humans is a premise for virus cross-species transmission, a more in depth investigation is needed to understand the infection capacity of these viruses in other animal cells, as well as to track and reveal its ecological dynamics in the wild, including biotic and abiotic factors that drive viral shedding and potential spillover events.

In addition to the deep ancestral inferences, high resolution phylogenetic reconstructions also allowed us to make inferences about the more recent *Ambecovirus* evolution. There are at least two clades circulating in the Americas: clade I, in sylvatic and sinantropic bats (*Pteronotus* spp. and *C. perspicillata*); and clade II, which circulates mainly in sinantropic bats (*Artibeus* spp. and *C. perspicillata*). Our findings also suggests that clade I strains infects mainly *Pteronotus* spp., but may be able to naturally infect multiple sympatric host species, i.e. species roosting in the same cave (*P. gymnonotus and C. perspicillata*, for example), or able to cross species barriers. Habitat overlap, and particularly host density at roostings with multiple bat species, have been shown to facilitate multispecies infection by coronaviruses due to close physical contact (76). However, more longitudinal data (e.g. shedding) from *P. gymnonotus* and *C. perspicillata* is needed in order to assess if our results represent an isolated event leading to virus extinction or if sustained transmission among *C. perspicillata* populations has been established. We were able to sample the same bat metapopulations in several time periods, but most of our sampling was restricted to the dry seasons, hence the temporal virus dynamics captured is limited for broad inferences about infection prevalence and shedding across time. In case sustained transmission is confirmed, it shall have two major implications: a) it will indicate cross-species transmission of *Betacoronavirus* in neotropical bats, contrary to the reports of neotropical bats-*Betacoronavirus* cospeciation and the implied resistance to host switching (40); and b) the detection of the new

ambecovirus in *C. perspicillata,* which is frequently found roosting in caves but is also a common synanthropic species, raises risk concerns, as this species is frequently found in large urban centers (77,78).

Interestingly, Anthony *et al*. (18) have pointed out that the absence of coronaviruses in some bats species was likely due to insufficient sampling. This is confirmed by several studies showing that only a third of the bat species were investigated for virus infections so far (73). In our study only *P. gymnonotus* had more than 200 specimens sampled, while an ideal sample size of more than 150-400 individuals is recommended for increasing the chances of detecting Coronaviruses in bats (18). However, population sizes for many of the sampled bat species is reduced with sometimes less than 25 individuals in a colony (i.e *Desmodus rotundus* at the Furna do Gato cave). Therefore, our sampling reflects relatively well the natural population sizes of these species. Still, the low sample size may have reduced the chance of detecting ambecoviruses in some of the studied bats. In addition, not only limited sampling contributes to such underestimation, but also the use of specific assays relying on previous information, particularly for highly divergent viruses such as ambecoviruses. Therefore, more frequent bat sampling and viral screening with unbiased methods are necessary to further characterize new coronaviruses and other zoonotic pathogens.

Perspectives

The discovery of the *Betacoronavirus* subgenus-*Ambecovirus* in neotropical bats, besides bringing new insights to viral evolution and ecology of these bats and viruses, raises some important questions: Are bats the main reservoir of this new subgenus? Does *Ambecovirus* infect only bats or also has the potential to infect other mammals, including humans? How frequently do these viruses cross the species barrier? While current evidence suggests host restriction to bat species (18,28,40) for Alphacoronaviruses and Betacoronaviruses of the *Merbecovirus* subgenus in the Americas, the detection of *Ambecovirus* in both sylvatic and synanthropic bat species raises concerns about potential cross-species transmission. Further understanding of the host range and specificity at molecular level also will strengthen our understanding of spike protein evolution, particularly of the S1-CDT region that might contain the RBD and RBM motifs. For this new CoV subgenus open questions also remain regarding the receptor used and the domain and residues that are more important for binding to the cell receptor ultimately defining tissue trophism, host range and potential epidemic potential. The recombination potential of Ambecoviruses also needs to be addressed in the future. Recombination is a well-documented driver of coronavirus evolution and emergence, as seen in SARS-CoV and MERS-CoV. However, due to the lack of closely related parental lineages and limited genomic data, recombination analysis for Ambecoviruses remains inconclusive. As more complete genomes are sequenced, it will be possible to assess recombination frequency and identify potential hotspots. Lastly, at the population

level, there is a need to further assess viral shedding and map the likelihood of host overlap with other wild and domestic species or humans to more accurately estimate the spillover and emergence risks. In summary, *Ambecovirus* represents a deeply divergent lineage within *Betacoronavirus* genus, and its discovery underscores the importance of continued virome exploration in under-sampled regions. Addressing these open questions will not only clarify the evolutionary trajectory of this subgenus but also inform risk assessments for future zoonotic events.

Data Availability

Supplementary materials are available at - Wallau, Gabriel (2025). Ambecovirus, a novel Betacoronavirus subgenus circulating in neotropical bats sheds new light on bat-borne coronaviruses evolution. figshare. Dataset. https://doi.org/10.6084/m9.figshare.29602244.v1.

Acknowledgements

We are very grateful to all the people who helped us during the fieldwork, especially Bárbara Coelho, Deibson Belo, Eduardo Pires, Jennifer Barros, Juliana Bezerra, Maria Júlia de França, and Narjara Pimentel. We also thank Heike Baum, Alexandra Bialonski, and Marike Petersen for their excellent technical assistance in metagenomic sequencing (mNGS).

Funding

The authors gratefully acknowledge support from the CAPES-Humboldt Research Fellowship. G.L.W. hold fellowships from Conselho Nacional de Desenvolvimento Científico e Tecnológico (Grant processes 307209/2023-7). E.Ba. holds a postdoctoral fellow at the São Paulo Research Foundation (FAPESP; grant #2023/09610-8).

References

1. Aziz RK, Breitbart M, Edwards RA. Transposases are the most abundant, most ubiquitous genes in nature. Nucleic Acids Res. 2010 Jul;38(13):4207–17.

2. Koonin EV, Krupovic M, Dolja VV. The global virome: How much diversity and how many independent origins? Environ Microbiol. 2023;25(1):40–4.

3. Van Brussel K, Holmes EC. Zoonotic disease and virome diversity in bats. Curr Opin Virol. 2022 Feb;52:192–202.

4. Jurburg SD, Hom EFY, Chatzinotas A. Beyond pathogenesis: Detecting the full spectrum of ecological interactions in the virosphere. PLOS Biol. 2023 May 15;21(5):e3002109.

5. Zhang YZ, Shi M, Holmes EC. Using Metagenomics to Characterize an Expanding Virosphere. Cell. 2018 Mar 8;172(6):1168–72.

6. Shi M, Lin XD, Tian JH, Chen LJ, Chen X, Li CX, et al. Redefining the invertebrate RNA virosphere. Nature. 2016 Dec 22;540(7634):539–43.

7. Råberg L, Graham AL, Read AF. Decomposing health: tolerance and resistance to parasites in animals. Philos Trans R Soc B Biol Sci. 2008 Oct 16;364(1513):37–49.

8. Irving AT, Ahn M, Goh G, Anderson DE, Wang LF. Lessons from the host defences of bats, a unique viral reservoir. Nature. 2021 Jan;589(7842):363–70.

9. Morales AE, Dong Y, Brown T, Baid K, Kontopoulos DG, Gonzalez V, et al. Bat genomes illuminate adaptations to viral tolerance and disease resistance. Nature. 2025 Feb;638(8050):449–58.

10. Ahn M, Anderson DE, Zhang Q, Tan CW, Lim BL, Luko K, et al. Dampened NLRP3-mediated inflammation in bats and implications for a special viral reservoir host. Nat Microbiol. 2019 May;4(5):789–99.

11. Solari S, Baker RJ. Mammal Species of the World: A Taxonomic and Geographic Reference by D. E. Wilson; D. M. Reeder. J Mammal. 2007 Jun 1;88(3):824–30.

12. Letko M, Seifert SN, Olival KJ, Plowright RK, Munster VJ. Bat-borne virus diversity, spillover and emergence. Nat Rev Microbiol. 2020 Aug;18(8):461–71.

13. Hernández-Aguilar I, Lorenzo C, Santos-Moreno A, Naranjo EJ, Navarrete-Gutiérrez D. Coronaviruses in Bats: A Review for the Americas. Viruses. 2021 Jul;13(7):1226.

14. Cohen LE, Fagre AC, Chen B, Carlson CJ, Becker DJ. Coronavirus sampling and surveillance in bats from 1996–2019: a systematic review and meta-analysis. Nat Microbiol. 2023 Jun;8(6):1176–86.

15. Vijaykrishna D, Smith GJD, Zhang JX, Peiris JSM, Chen H, Guan Y. Evolutionary insights into the ecology of coronaviruses. J Virol. 2007 Apr;81(8):4012–20.

16. Cui J, Li F, Shi ZL. Origin and evolution of pathogenic coronaviruses. Nat Rev Microbiol. 2019 Mar;17(3):181–92.

17. Woo PCY, Lau SKP, Lam CSF, Lau CCY, Tsang AKL, Lau JHN, et al. Discovery of Seven Novel Mammalian and Avian Coronaviruses in the Genus Deltacoronavirus Supports Bat Coronaviruses as the Gene Source of Alphacoronavirus and Betacoronavirus and Avian Coronaviruses as the Gene

Source of Gammacoronavirus and Deltacoronavirus. J Virol. 2012 Apr;86(7):3995–4008.

18. Anthony SJ, Johnson CK, Greig DJ, Kramer S, Che X, Wells H, et al. Global patterns in coronavirus diversity. Virus Evol. 2017 Jan;3(1):vex012.

19. Wallau GL, Barbier E, Tomazatos A, Schmidt-Chanasit J, Bernard E. The virome of bats inhabiting Brazilian biomes: knowledge gaps and biases towards zoonotic viruses [Internet]. bioRxiv; 2022 [cited 2023 Jan 8]. p. 2022.10.11.511773. Available from: https://www.biorxiv.org/content/10.1101/2022.10.11.511773v1

20. Zhang YZ, Chen YM, Wang W, Qin XC, Holmes EC. Expanding the RNA Virosphere by Unbiased Metagenomics. Annu Rev Virol. 2019 Sep 29;6(1):119–39.

21. Abranches S. Biological Megadiversity as a Tool of Soft Power and Development for Brazil. Braz Polit Sci Rev. 2020 Aug 17;14:e0006.

22. Carlson CJ, Albery GF, Merow C, Trisos CH, Zipfel CM, Eskew EA, et al. Climate change increases cross-species viral transmission risk. Nature. 2022 Apr 28;1–1.

23. Olival KJ, Hosseini PR, Zambrana-Torrelio C, Ross N, Bogich TL, Daszak P. Host and viral traits predict zoonotic spillover from mammals. Nature. 2017 Jun;546(7660):646–50.

24. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. Nature. 2008;451(7181):990–3.

25. Drexler JF, Corman VM, Drosten C. Ecology, evolution and classification of bat coronaviruses in the aftermath of SARS. Antiviral Res. 2014 Jan;101:45–56.

26. Simmons NB, Cirranello AL. Bat Species of the World: A Taxonomic and Geographical Database. 2024.

27. Carrington CVF, Foster JE, Zhu HC, Zhang JX, Smith GJD, Thompson N, et al. Detection and phylogenetic analysis of group 1 coronaviruses in South American bats. Emerg Infect Dis. 2008 Dec;14(12):1890–3.

28. Caraballo DA, Sabio MS, Colombo VC, Piccirilli MG, Vico L, Hirmas Riade SM, et al. The Role of Molossidae and Vespertilionidae in Shaping the Diversity of Alphacoronaviruses in the Americas. Microbiol Spectr. 2022 Oct 12;10(6):e03143-22.

29. Cerri A, Bolatti EM, Zorec TM, Montani ME, Rimondi A, Hosnjak L, et al. 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. Microbiol Spectr. 2023 Sep 11;0(0):e02047-23.

30. Bergner LM, Orton RJ, Streicker DG. Complete Genome Sequence of an Alphacoronavirus from Common Vampire Bats in Peru. Microbiol Resour Announc. 2020 Aug 20;9(34):e00742-20.

31. Alves RS, do Canto Olegário J, Weber MN, da Silva MS, Canova R, Sauthier JT, et al. Detection of coronavirus in vampire bats (Desmodus rotundus) in southern Brazil. Transbound Emerg Dis. 2021 May 12;

32. Bittar C, Machado RRG, Comelis MT, Bueno LM, Beguelini MR, Morielle-Versute E, et al. Alphacoronavirus Detection in Lungs, Liver, and Intestines of Bats from Brazil. Microb Ecol. 2020 Jan;79(1):203–12.

33. Góes LGB, Campos AC de A, Carvalho C de, Ambar G, Queiroz LH, Cruz-Neto AP, et al.

Genetic diversity of bats coronaviruses in the Atlantic Forest hotspot biome, Brazil. Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis. 2016 Oct;44:510–3.

34. Lima FE de S, Campos FS, Kunert Filho HC, Batista HB de CR, Carnielli P, Cibulski SP, et al. Detection of Alphacoronavirus in velvety free-tailed bats (Molossus molossus) and Brazilian free-tailed bats (Tadarida brasiliensis) from urban area of Southern Brazil. Virus Genes. 2013 Aug;47(1):164–7.

35. Brandão PE, Scheffer K, Villarreal LY, Achkar S, Oliveira R de N, Fahl W de O, et al. A coronavirus detected in the vampire bat Desmodus rotundus. Braz J Infect Dis Off Publ Braz Soc Infect Dis. 2008 Dec;12(6):466–8.

36. Simas PVM, Barnabé AC de S, Durães-Carvalho R, Neto DF de L, Caserta LC, Artacho L, et al. Bat coronavirus in Brazil related to appalachian ridge and porcine epidemic diarrhea viruses. Emerg Infect Dis. 2015 Apr;21(4):729–31.

37. Anthony SJ, Ojeda-Flores R, Rico-Chávez O, Navarrete-Macias I, Zambrana-Torrelio CM, Rostal MK, et al. Coronaviruses in bats from Mexico. J Gen Virol. 2013 May;94(Pt 5):1028–38.

38. Bueno LM, Rizotto LS, Viana A de O, Silva LMN, de Moraes MV dos S, Benassi JC, et al. High genetic diversity of alphacoronaviruses in bat species (Mammalia: Chiroptera) from the Atlantic Forest in Brazil. Transbound Emerg Dis [Internet]. [cited 2022 Jul 20];n/a(n/a). Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/tbed.14636

39. Figueiroa T, Galvão Bueno M, Bento Moura PE, de Oliveira MB, Passos Cordeiro JL, Santos-Cavalcante N, et al. Alpha and Betacoronavirus Detection in Neotropical Bats from Northeast Brazil Suggests Wide Geographical Distribution and Persistence in Natural Populations. Anim Open Access J MDPI. 2025 Jan 24;15(3):332.

40. Caraballo DA. Cross-Species Transmission of Bat Coronaviruses in the Americas: Contrasting Patterns between Alphacoronavirus and Betacoronavirus. Microbiol Spectr. 2022 Jun 30;10(4):e01411-22.

41. Silvério BS, Guilardi MD, Martins JO, Duro RLS, de Sousa LLF, Cabral-Miranda G, et al. Coronavirus Cryptic Landscape and Draft Genome of a Novel CoV Clade Related to MERS From Bats Circulating in Northeastern Brazil. J Med Virol. 2025;97(1):e70173.

42. Díaz MM, Aguirre L, Barquez R. Clave de identificación de los murciélagos del cono sur de Sudamérica. Centro de Estudios en Biología Teórica y Aplicada, Cochabamba.; 2011.

43. Díaz MM, Solari S, Gregorin R, Aguirre LF, Barquez RM. Clave de Identificación de los Murciélagos Neotropicales. Tucumán: Fundación PCMA; 2021. 211 p.

44. Gardner AL. Mammals of South America. III. Vol. 1. Chicago: University of Chicago Press; 2008.

45. Pavan AC. Family Mormoopidae (ghost-faced, naked-backed and mustached bats). In: Handbook of the Mammals of the World: Bats. Barcelona: Lynx Edicions; 2019. p. 424–43.

46. Kunz TH, Anthony ELP. Age Estimation and Post-Natal Growth in the Bat Myotis lucifugus. J Mammal. 1982 Feb 25;63(1):23–32.

47. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinforma Oxf Engl. 2014 Aug 1;30(15):2114–20.

48. Bushnell B. Archive. 2014 [cited 2025 Jul 10]. BBTools software package. Available from: https://archive.jgi.doe.gov/data-and-tools/software-tools/bbtools/

49. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013 Jan;41(Database issue):D590-596.

50. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015 May 15;31(10):1674–6.

51. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015 Jan;12(1):59–60.

52. da Silva AF, da Silva Neto AM, Aksenen CF, Jeronimo PMC, Dezordi FZ, Almeida SP, et al. ViralFlow v1.0—a computational workflow for streamlining viral genomic surveillance. NAR Genomics Bioinforma. 2024 Jun 1;6(2):lqae056.

53. Stn A, Rjp N, Jn N, Ap C, Gw T, Bj W. CoverM: read alignment statistics for metagenomics. Bioinforma Oxf Engl [Internet]. 2025 Mar 29 [cited 2025 Jul 4];41(4). Available from: https://pubmed.ncbi.nlm.nih.gov/40193404/

54. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010 Mar 8;11:119.

55. Huson DH, Beier S, Flade I, Górska A, El-Hadidi M, Mitra S, et al. MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. PLoS Comput Biol. 2016 Jun;12(6):e1004957.

56. Gouilh MA, Puechmaille SJ, Gonzalez JP, Teeling E, Kittayapong P, Manuguerra JC. SARS-Coronavirus ancestor's foot-prints in South-East Asian bat colonies and the refuge theory. Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis. 2011 Oct;11(7):1690–702.

57. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform. 2019 Jul 19;20(4):1160–6.

58. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics. 2014 Nov 15;30(22):3276–8.

59. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol. 2020 May 1;37(5):1530–4.

60. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017 Jun;14(6):587–9.

61. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. 2018 Jan;4(1):vey016.

62. Papadopoulos JS, Agarwala R. COBALT: constraint-based alignment tool for multiple protein sequences. Bioinforma Oxf Engl. 2007 May 1;23(9):1073–9.

63. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021 Aug;596(7873):583–9.

64. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 2018 Jul 2;46(W1):W296–303.

65. Jurrus E, Engel D, Star K, Monson K, Brandi J, Felberg LE, et al. Improvements to the APBS biomolecular solvation software suite. Protein Sci. 2018;27(1):112–28.

66. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. Nucleic Acids Res. 2004 Jul 1;32(suppl_2):W665–7.

67. Jin M, Hassan Z, Li Z, Liu Y, Marakhovskaia A, Wong AHM, et al. Human coronavirus HKU1 spike structures reveal the basis for sialoglycan specificity and carbohydrate-promoted conformational changes. Nat Commun. 2025 May 5;16(1):4158.

68. Ji W, Peng Q, Fang X, Li Z, Li Y, Xu C, et al. Structures of a deltacoronavirus spike protein bound to porcine and human receptors. Nat Commun. 2022 Mar 18;13(1):1467.

69. Naveca FG, Nascimento V, Souza V, Corado A de L, Nascimento F, Silva G, et al. Spread of Gamma (P.1) Sub-Lineages Carrying Spike Mutations Close to the Furin Cleavage Site and Deletions in the N-Terminal Domain Drives Ongoing Transmission of SARS-CoV-2 in Amazonas, Brazil. Microbiol Spectr. 2022 Feb 23;10(1):e02366-21.

70. Ferraz MVF, Moreira EG, Coêlho DF, Wallau GL, Lins RD. Immune evasion of SARS-CoV-2 variants of concern is driven by low affinity to neutralizing antibodies. Chem Commun. 2021 Jun 17;57(49):6094–7.

71. Dhamotharan K, Korn SM, Wacker A, Becker MA, Günther S, Schwalbe H, et al. A core network in the SARS-CoV-2 nucleocapsid NTD mediates structural integrity and selective RNA-binding. Nat Commun. 2024 Dec 9;15(1):10656.

72. Forni D, Cagliani R, Clerici M, Sironi M. Molecular Evolution of Human Coronavirus Genomes. Trends Microbiol. 2017 Jan 1;25(1):35–48.

73. Ruiz-Aravena M, McKee C, Gamble A, Lunn T, Morris A, Snedden CE, et al. Ecology, evolution and spillover of coronaviruses from bats. Nat Rev Microbiol. 2021 Nov 19;1–16.

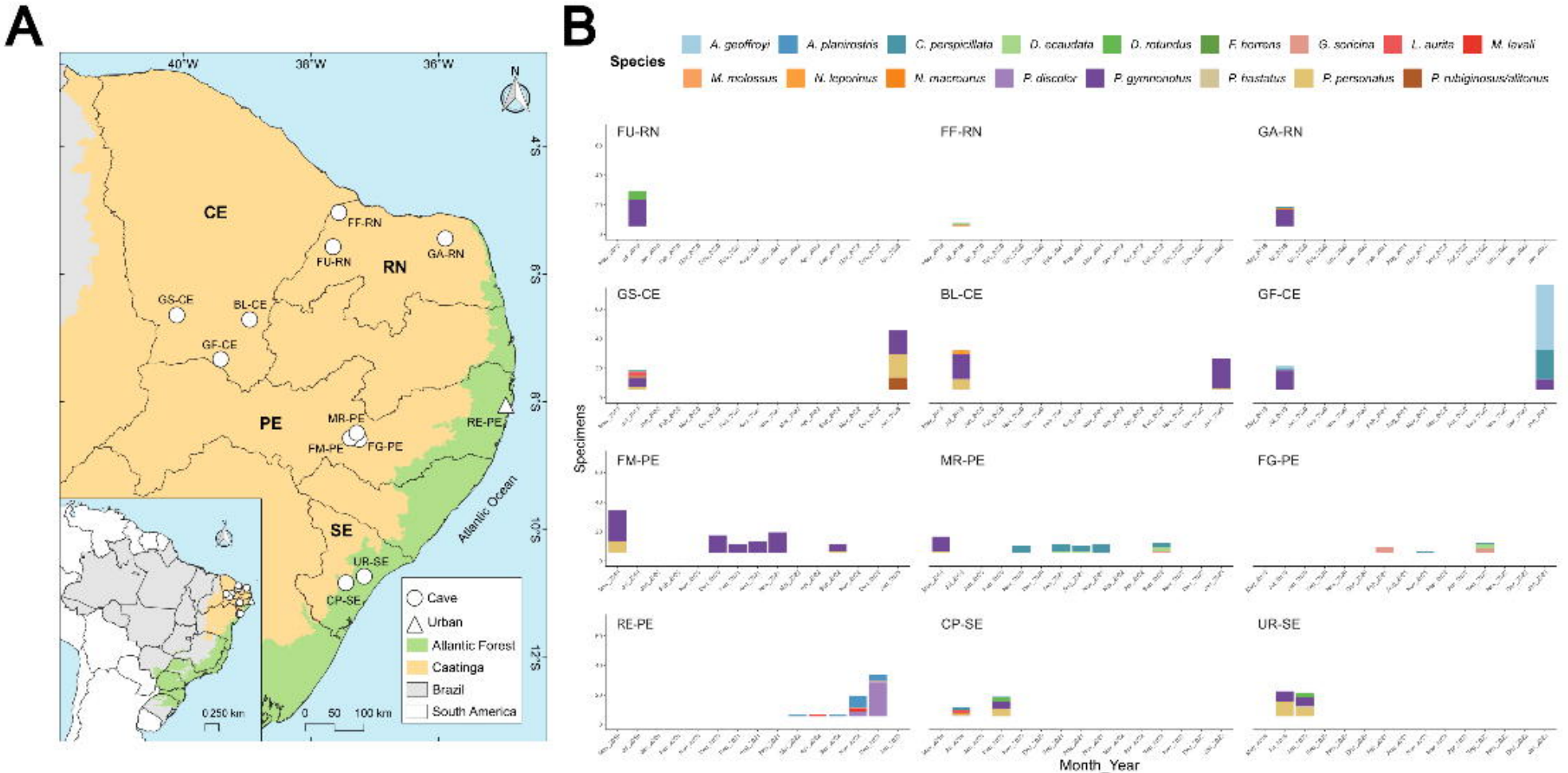
74. Li F. Structure, Function, and Evolution of Coronavirus Spike Proteins. Annu Rev Virol. 2016 Sep 29;3(1):237–61.

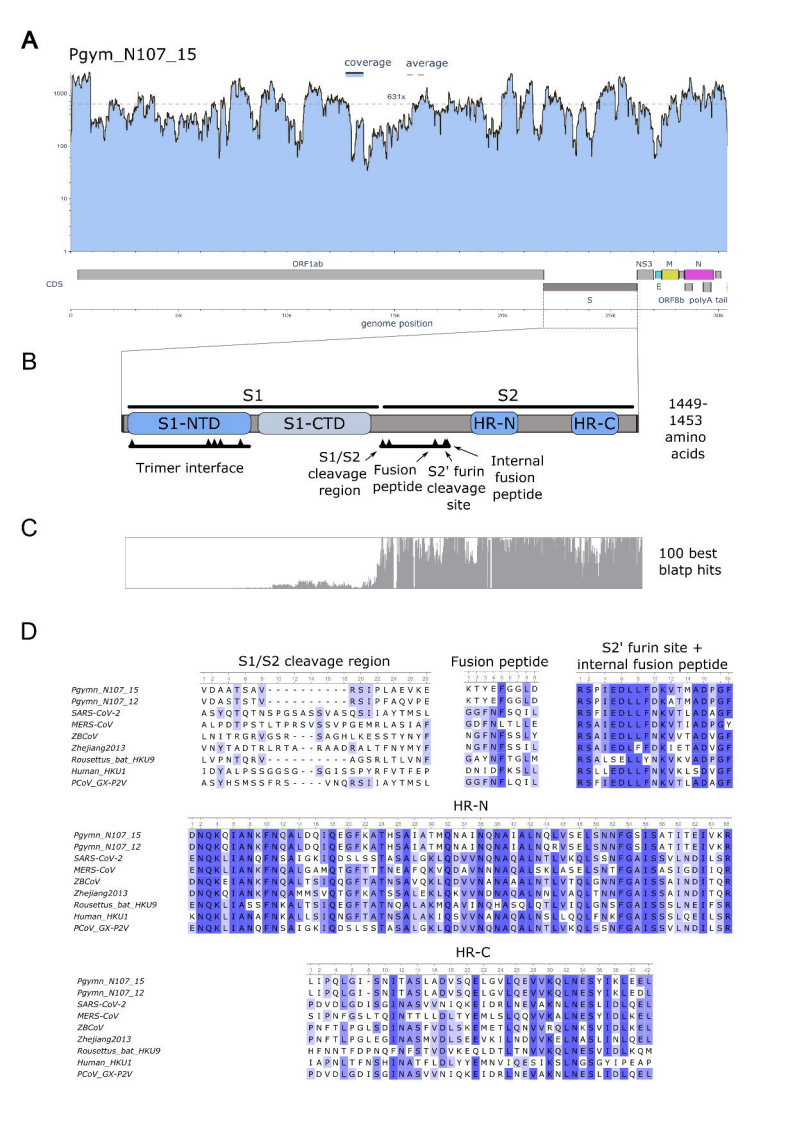
75. Li F. Evidence for a Common Evolutionary Origin of Coronavirus Spike Protein Receptor-Binding Subunits. J Virol. 2012 Mar;86(5):2856–8.

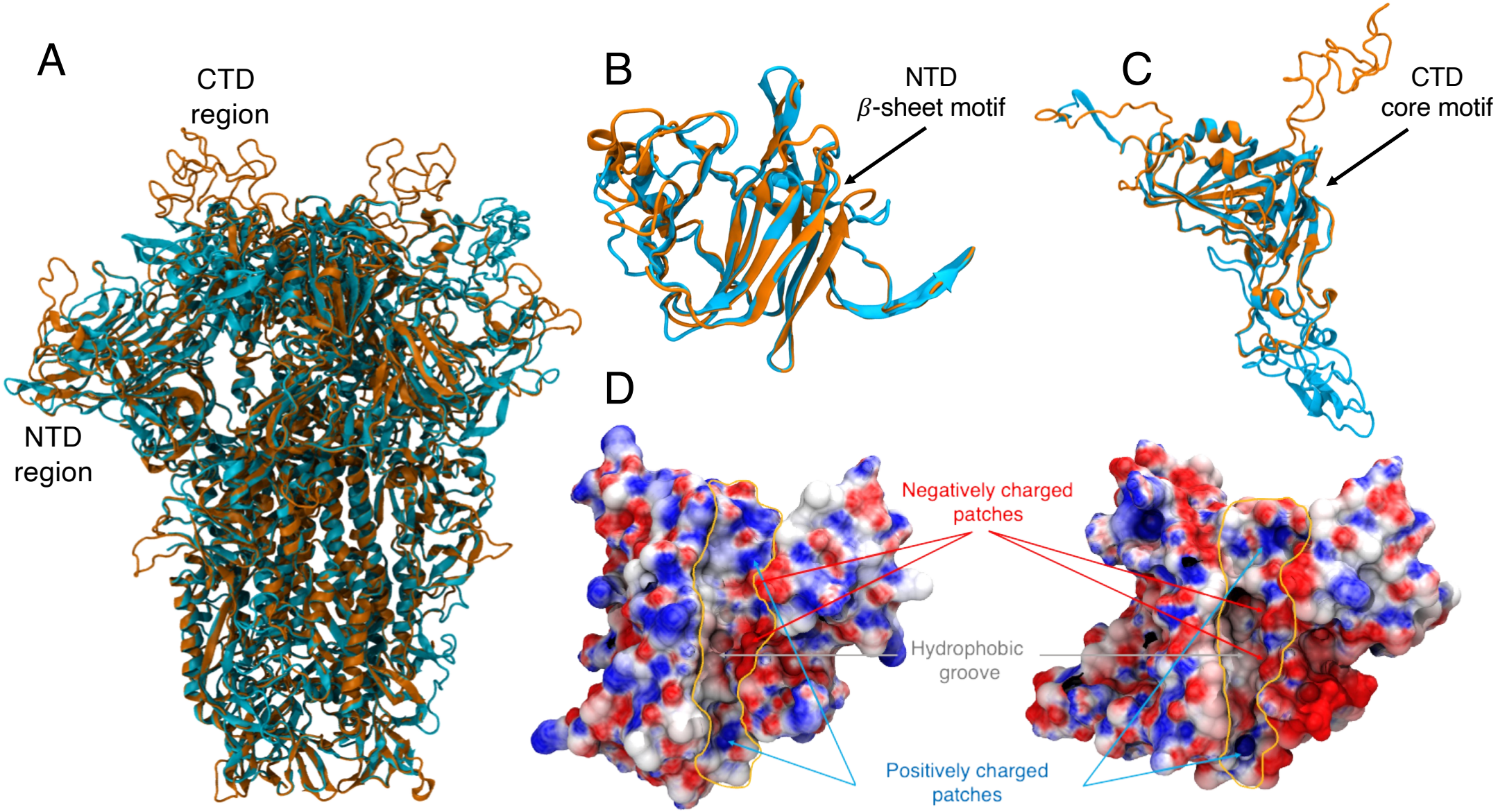
76. Willoughby AR, Phelps KL, PREDICT Consortium, Olival KJ. A Comparative Analysis of Viral Richness and Viral Sharing in Cave-Roosting Bats. Diversity. 2017 Sep;9(3):35.

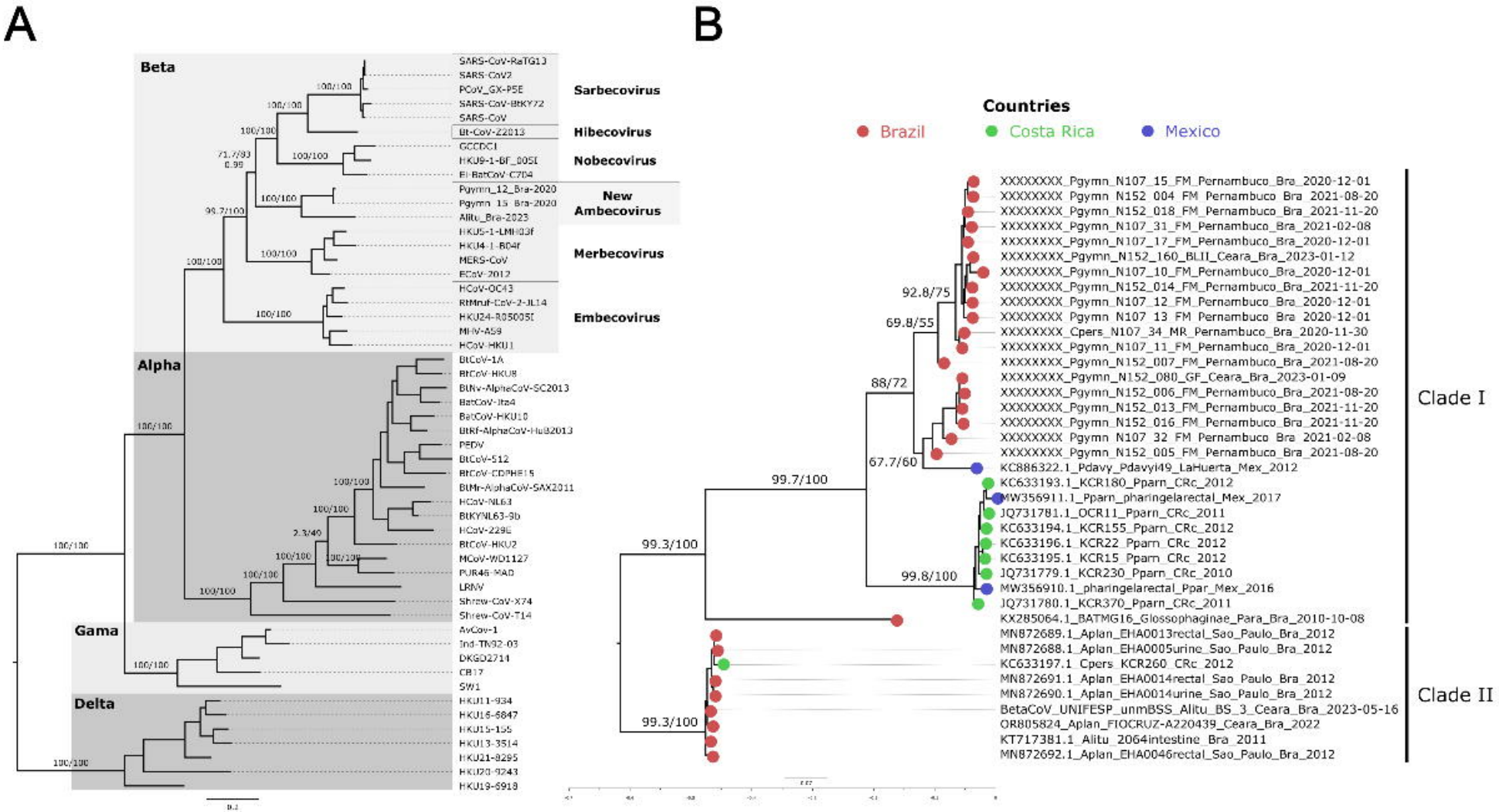
77. Barros JS, Bernard E. Species richness, occurrence and rarity of bats in Brazilian caves. Austral Ecol. 2023;48(8):2144–70.

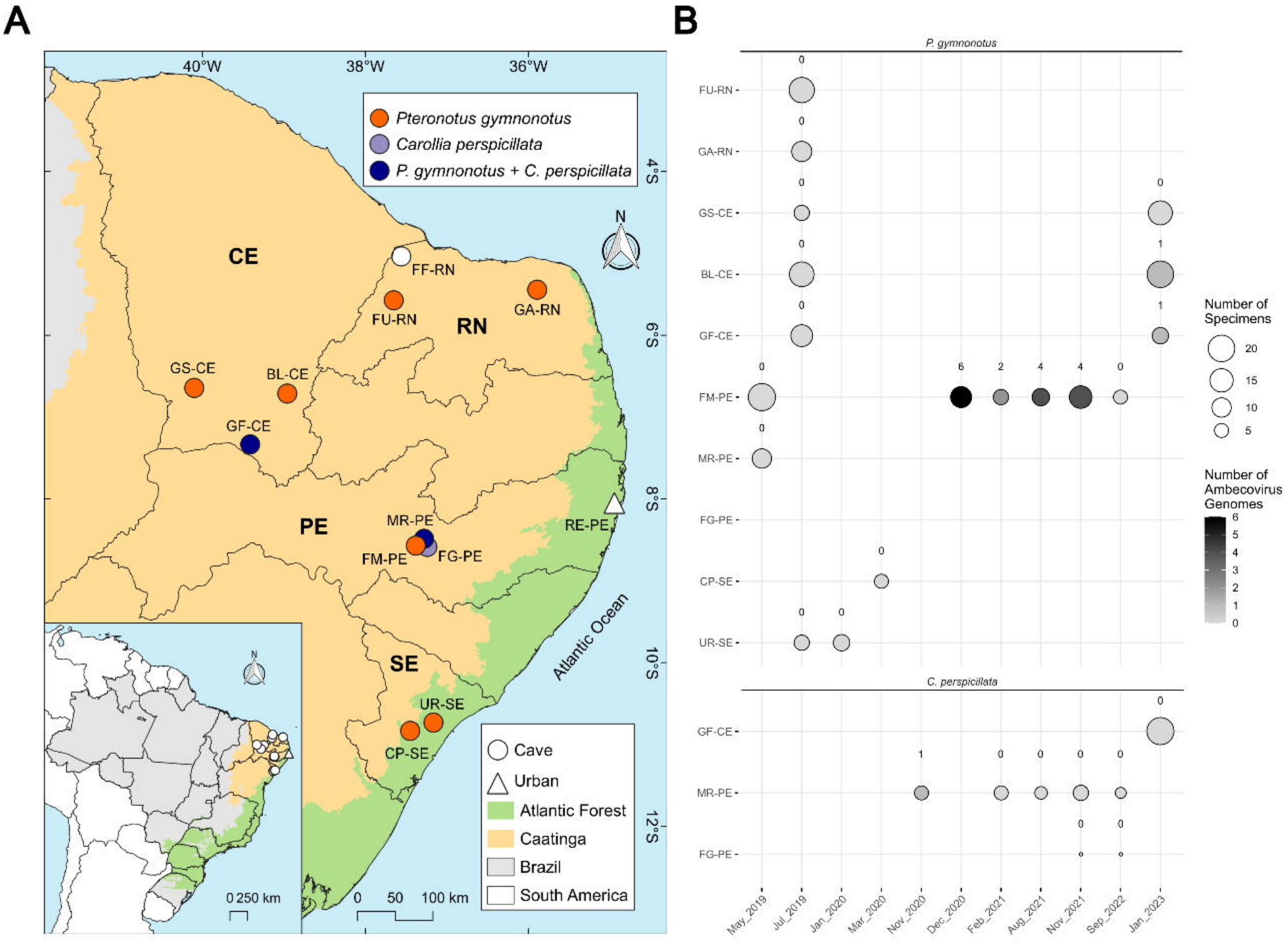
78. Nunes H, Rocha FL, Cordeiro-Estrela P. Bats in urban areas of Brazil: roosts, food resources and parasites in disturbed environments. Urban Ecosyst. 2017 Aug 1;20(4):953–69.


A
40°W
38°W
36°W
N
4°S
6°S
8°S
10°S
12°S
CE
RN
PE
SE
FF-RN
FU-RN
GA-RN
GS-CE
BL-CE
GF-CE
MR-PE
FM-PE
FG-PE
RE-PE
UR-SE
CP-SE
Atlantic Ocean
Cave
Urban
Atlantic Forest
Caatinga
Brazil
South America
0 250 km
0 50 100 km
B
Species
A. geoffroyi
A. planirostris
C. perspicillata
D. ecaudata
D. rotundus
E. bombus
G. soricina
L. aurita
M. levali
M. molossus
N. leporinus
N. macrosus
P. discolor
P. gymnonotus
P. hastatus
P. personatus
P. rubiginosus/albiventus
FU-RN
FF-RN
GA-RN
GS-CE
BL-CE
GF-CE
FM-PE
MR-PE
FG-PE
RE-PE
CP-SE
UR-SE
Specimens
Month_Year


A
Pgym_N107_15
coverage
average
631x
1000
100
10
1
DRF1ab
CDS
NS3
M
N
E
S
ORF8b
polyA tail
0
5K
10K
15K
20K
25K
30K
genome position
B
S1
S2
S1-NTD
S1-CTD
HR-N
HR-C
1449-1453 amino acids
Trimer interface
S1/S2 cleavage region
Fusion peptide
S2' furin cleavage site
Internal fusion peptide
C
100 best blatp hits
D
S1/S2 cleavage region
Fusion peptide
S2' furin site + internal fusion peptide
Pgymn_N107_15
Pgymn_N107_12
SARS-CoV-2
MERS-CoV
ZBCoV
Zhejiang2013
Rousettus_bat_HKU9
Human_HKU1
PCoV_GX-P2V
HR-N
HR-C


A
CTD
region
NTD
region
B
NTD
β-sheet motif
C
CTD
core motif
D
Negatively charged
patches
Hydrophobic
groove
Positively charged
patches


A
Beta
Alpha
Gama
Delta
Sarbecovirus
Hibecovirus
Nobecovirus
New Ambecovirus
Merbecovirus
Embecovirus
SARS-CoV-RaTG13
SARS-CoV2
PCoV_GX-P5E
SARS-CoV-BtKY72
SARS-CoV
Bt-CoV-Z2013
GCCDCI
HKU9-1-BF_005E
EI-BatCoV-C704
Pgymn_12_Bra-2020
Pgymn_15_Bra-2020
Alitu_Bra-2023
HKU5-1-LMH03f
HKU4-1-BD4f
MERS-CoV
ECoV-2012
HCoV-OC43
RtMruf-CoV-2-JL14
HKU24-R05003E
MHV-A59
HCoV-HKU1
BtCoV-1A
BtCoV-HKU8
BtNv-AlphaCoV-SC2013
BatCoV-Itz4
BatCoV-HKU10
BtRf-AlphaCoV-Hu2013
PEDV
BtCoV-512
BtCoV-CDPHE15
BtMr-AlphaCoV-SAX2011
HCoV-NL63
BtKYNL63-9b
HCoV-229E
BtCoV-HKU2
MCoV-WD1127
PUR46-MAD
LRMV
Shrew-CoV-X74
Shrew-CoV-T14
AvCov-1
Ind-TN92-03
DKGD2714
CB17
SW1
HKU11-934
HKU16-6847
HKU15-155
HKU13-3514
HKU21-8295
HKU20-9243
HKU19-6918
B
Countries
Brazil
Costa Rica
Mexico
Clade I
Clade II
XXXXXXXX_Pgymn_N107_15_FM_Pernambuco_Bra_2020-12-01
XXXXXXXX_Pgymn_N152_004_FM_Pernambuco_Bra_2021-08-20
XXXXXXXX_Pgymn_N152_018_FM_Pernambuco_Bra_2021-11-20
XXXXXXXX_Pgymn_N107_31_FM_Pernambuco_Bra_2021-02-08
XXXXXXXX_Pgymn_N107_17_FM_Pernambuco_Bra_2020-12-01
XXXXXXXX_Pgymn_N152_160_BLII_Ceara_Bra_2023-01-12
XXXXXXXX_Pgymn_N107_10_FM_Pernambuco_Bra_2020-12-01
XXXXXXXX_Pgymn_N152_014_FM_Pernambuco_Bra_2021-11-20
XXXXXXXX_Pgymn_N107_12_FM_Pernambuco_Bra_2020-12-01
XXXXXXXX_Pgymn_N107_13_FM_Pernambuco_Bra_2020-12-01
XXXXXXXX_Cpers_N107_34_MR_Pernambuco_Bra_2020-11-30
XXXXXXXX_Pgymn_N107_11_FM_Pernambuco_Bra_2020-12-01
XXXXXXXX_Pgymn_N152_007_FM_Pernambuco_Bra_2021-08-20
XXXXXXXX_Pgymn_N152_080_GF_Ceara_Bra_2023-01-09
XXXXXXXX_Pgymn_N152_006_FM_Pernambuco_Bra_2021-08-20
XXXXXXXX_Pgymn_N152_013_FM_Pernambuco_Bra_2021-11-20
XXXXXXXX_Pgymn_N152_016_FM_Pernambuco_Bra_2021-11-20
XXXXXXXX_Pgymn_N107_32_FM_Pernambuco_Bra_2021-02-08
XXXXXXXX_Pgymn_N152_005_FM_Pernambuco_Bra_2021-08-20
KC886322.1_Pdavy_Pdavyi49_LaHuerta_Mex_2012
KC633193.1_KCR180_Pparn_CRc_2012
MW356911.1_Pparn_pharingelarectal_Mex_2017
JQ731781.1_OCR11_Pparn_CRc_2011
KC633194.1_KCR155_Pparn_CRc_2012
KC633196.1_KCR22_Pparn_CRc_2012
KC633195.1_KCR15_Pparn_CRc_2012
JQ731779.1_KCR230_Pparn_CRc_2010
MW356910.1_pharingelarectal_Ppar_Mex_2016
JQ731780.1_KCR370_Pparn_CRc_2011
KX285064.1_BATMG16_Glossophaginae_Para_Bra_2010-10-08
MN872689.1_Aplan_EHA0013rectal_Sao_Paulo_Bra_2012
MN872688.1_Aplan_EHA0005urine_Sao_Paulo_Bra_2012
KC633197.1_Cpers_KCR260_CRc_2012
MN872691.1_Aplan_EHA0014rectal_Sao_Paulo_Bra_2012
MN872690.1_Aplan_EHA0014urine_Sao_Paulo_Bra_2012
BetaCoV_UNIFESP_unmBSS_Alitu_BS_3_Ceara_Bra_2023-05-16
OR805824_Aplan_FIOCRUZ-A220439_Ceara_Bra_2022
KT717381.1_Alitu_2064intestine_Bra_2011
MN872692.1_Aplan_EHA0046rectal_Sao_Paulo_Bra_2012
92.8/75
69.8/55
88/72
67.7/60
99.7/100
99.3/100
99.8/100
99.3/100


A
40°W
38°W
36°W
4°S
6°S
8°S
10°S
12°S
Pteronotus gymnonotus
Carollia perspicillata
P. gymnonotus + C. perspicillata
CE
RN
PE
SE
FF-RN
FU-RN
GA-RN
GS-CE
BL-CE
GF-CE
MR-PE
FM-PE
FG-PE
RE-PE
UR-SE
CP-SE
Atlantic Ocean
N
Cave
Urban
Atlantic Forest
Caatinga
Brazil
South America
0 250 km
0 50 100 km
B
P. gymnonotus
C. perspicillata
FU-RN
GA-RN
GS-CE
BL-CE
GF-CE
FM-PE
MR-PE
FG-PE
CP-SE
UR-SE
May_2019
Jul_2019
Jan_2020
Mar_2020
Nov_2020
Dec_2020
Feb_2021
Aug_2021
Nov_2021
Sep_2022
Jan_2023
Number of Specimens
20
15
10
5
Number of Ambecovirus Genomes
6
5
4
3
2
1
0