**Title**

Full genome sequences of novel *Nobecoviruses* identified in endemic Madagascar fruit bats

**Authors**

Gwenddolen Kettenburg1\*, Amy Kistler2, Vida Ahyong2, Angelo Andrianiaina3, Santino Andry4, Joseph L. DeRisi2, Anecia Gentles5, Hafaliana Christian Ranaivoson3,6, ﻿Tsiry Hasina Randriambolamanantsoa6, Ny Anjara Fifi Ravelomanantsoa3, Cristina M. Tato2, Jean-Michel Héraud6, Philippe Dussart6, and Cara E. Brook1

**Author affiliations**

1Department of Ecology and Evolution, University of Chicago, Chicago, IL, U.S.A.

2Chan Zuckerberg Biohub, San Francisco, CA, U.S.A.

3Department of Zoology and Animal Biodiversity, University of Antananarivo, Antananarivo, Madagascar.

4Department of Entomology, University of Antananarivo, Antananarivo, Madagascar.

5Odum School of Ecology, University of Georgia, Athens, GA, U.S.A.

6Virology Unit, Institut Pasteur of Madagascar, Antananarivo, Madagascar.

\*Corresponding author, [gkettenburg@uchicago.edu](mailto:gkettenburg@uchicago.edu)

**Keywords:**

*Nobecovirus, bat-borne coronavirus, recombination, zoonosis, Madagascar*

**Abstract**

Bats are natural reservoirs for both *Alpha*- and *Betacoronaviruses* and the hypothesized original hosts of five of seven known zoonotic coronaviruses. To date, the vast majority of bat coronavirus research has been concentrated in Asia, though coronaviruses are globally distributed; indeed, SARS-CoV and SARS-CoV-2-related *Betacoronaviruses* in the subgenus Sarbecovirus have been identified circulating in *Rhinolophid* bats in both Africa and Europe, despite the relative dearth of surveillance in these regions. In part with a long-term study examining the dynamics of potentially zoonotic viruses in three species of endemic Madagascar fruit bat (*Pteropus rufus, Eidolon dupreanum, Rousettus madagascariensis*), we carried out metagenomic Next Generation Sequencing on urine, throat, and fecal samples obtained from wild-caught individuals. We report detection of RNA derived from *Betacoronavirus* subgenus *Nobecovirus* in fecal samples from all three species and describe full genome sequences of novel *Nobecoviruses* in *P. rufus* and *R. madagascariensis.* Phylogenetic analysis indicates the existence of five distinct *Nobecovirus* clades, one of which is defined by the highly divergent sequence reported here from *P. rufus* bats*.* Madagascar *Nobecoviruses* derived from *P. rufus* and *R. madagascariensis* demonstrate, respectively, Asian and African phylogeographic origins, mirroring those of their fruit bat hosts. Bootscan recombination analysis indicates significant genomic reassortment has taken place in the spike, nucleocapsid, and NS7 accessory protein regions of the genome for viruses derived from both bat hosts. Madagascar offers a unique phylogeographic nexus of bats and viruses with both Asian and African phylogeographic origins, offering opportunities for unprecedented mixing of viral groups and, potentially, unique avenues for recombination. As fruit bats are consumed widely across Madagascar for subsistence, understanding the landscape of potentially zoonotic coronavirus circulation is essential to mitigation of future zoonotic threats.

**Introduction**

In the past 20 years, bat-derived coronaviruses SARS-CoV, MERS-CoV, and SARS-CoV-2 have been responsible for two deadly epidemics and the ongoing COVID-19 pandemic (1–4). These coronaviruses (CoVs) are members of the *Betacoronavirus* genus, which, along with genus *Alphacoronavirus*, are primarily associated with bat hosts (1–4); the remaining CoV genera, *Gammacoronavirus and Deltacoronavirus,* are typically hosted by birds (5). The *Betacoronavirus* group can be further broken down into bat-associated subgenera *Sarbecovirus* (hosted by bats in family Rhinolophidae (6,7)), *Merbecovirus* (hosted by bats in family Vespertilionidae(8–10)), *Nobecovirus* (hosted by bats in family Pteropodidae (11–13)), and *Hibecovirus* (hosted by bats in family Hipposideridae (14,15). The final *Betacoronavirus* subgenus, *Embecovirus,* is primarily associated with rodent and bovid hosts (16,17). Since the emergence of SARS-CoV in 2002, there has been increasing interest in surveying potential hosts of coronaviruses and contributing new virus sequences to public databases, with most effort focused on sampling bats from Asia (18–25), the continent of origin for both the SARS-CoV epidemic and the SARS-CoV-2 pandemic. More recently, there has arisen a more concerted effort to survey the landscape of bat-borne coronaviruses in other regions, including Africa and Europe (11,13,26–30).

The family Coronaviridae is considered one of the most likely viral taxa to switch host species (31,32), in part because many CoVs utilize well-conserved cell surface receptors to gain entry into a wide variety of mammalian host cells. The zoonotic Sarbecoviruses, SARS-CoV and SARS-CoV-2, for example, use the human cell surface receptor Angiotensin-converting enzyme 2 (ACE2) to gain entry into human cells (33,34), while many Merbecoviruses interact with the well-conserved vertebrate host cell receptor ﻿dipeptidyl peptidase 4 (DPP4) to do the same (35). Because CoVs are notoriously inclined towards recombination (with other CoVs, or more rarely, with other viral groups) there is concern that naturally-circulating CoVs presently unable to infect humans may acquire this ability in the future (26,38–41), a concern which highlights the importance of widespread CoV surveillance aimed at characterizing the global landscape of potentially CoV diversity available for recombination. Several factors, which have been reviewed at length elsewhere (31,42,43), contribute to the CoV affinity for recombination, including the large CoV genome size supported by a unique proofreading mechanism in the RNA-dependent RNA polymerase (RdRp) (44–47), as well as a ‘copy choice’ template switching mechanism of RNA replication whereby RdRp physically detaches from one RNA template during replication and reattaches to an adjacent template, thus facilitating recombination in cases where multiple viruses may be coinfecting the same cell (48).

Madagascar is an island country in southeastern Sub-Saharan Africa, located in the Indian Ocean, ~400 km off the coast from Mozambique. Madagascar has been isolated from the African continent for over 170 million years and all surrounding landmasses for over 80 million years, allowing for the evolution of a unique and highly endemic floral and faunal assemblage (49). The country is home to 51 species of bat (50), some three-quarters of which are endemic and boast long evolutionary divergence times with sister species on both the African and Asian continents (51–53). A growing body of work has characterized the landscape of potentially zoonotic viruses in Madagascar bats, identifying evidence of circulating infection (through RNA detection or serology) with henipaviruses, filoviruses, lyssaviruses, and coronaviruses (29,54–56). Previously coronavirus surveillance efforts have identified *Alphacoronavirus* RNA in the Malagasy insectivorous bat, *Mormopterus jugalaris,* and *Betacoronavirus* RNA in all three endemic Malagasy fruit bat species, *Pteropus rufus, Eidolon dupreanum,* and *Rousettus madagascariensis* (62,100). Previous studies have demonstrated that this latter *Betacoronavirus* RNA clusters with subgenus *Nobecovirus* (29,55); *Nobecoviruses* have been previously described infecting Pteropodidae fruit bats across Asia and in both East and West Africa (21,28,57–60). Though Nobecoviruses are not known to be zoonotic, previous research has described widespread circulation of a recombinant Nobecovirus carrying an orthoreovirus insertion throughout Asia (21,60,61), highlighting the capacity for this viral subgenus to undertake rapid shifts in genomic organization which could lead to expanded host range. As both *Eidolon dupreanum* and *Rousettus madagascariensis* are known to co-roost with each other, and with several species of insectivorous bat (62), CoV recombination is a distinct concern in the Madagascar system. Though no *Rhinolophus* spp. bats, the typical host for ACE2-using *Sarbecoviruses*, inhabit Madagascar, the island is home to four species of Hipposiderid bat, which host the *Sarbecovirus-*adjacent and understudied *Hibecoviruses,* as well as several species of Vespertilionid bat, the most common hosts for the zoonotic *Merbecoviruses*.

Human-bat contact rates are high in Madagascar, where bats are consumed widely as a source of human food and frequently roost in close proximity to human settlements or natural tourist attractions (63–66). In addition to the natural CoV diversity circulating in Malagasy bats, several human coronaviruses are known to circulate widely in Madagascar, including the common cold-causing Embecoviruses, HCoV-OC43 and HCoV-HKU1, and, more recently, the zoonotic Sarbecovirus, SARS-CoV-2 (67–69). As spillback of SARS-CoV-2 into wildlife hosts and possible recombination with wildlife viruses remains a global concern (70), characterization of the genetic diversity of bat-borne coronaviruses in Madagascar and elsewhere in Africa is a critical public health priority. Here we contribute and characterize three full genome sequences of two novel *Nobecoviruses*, derived *R. madagascariensis* and *P. rufus* hosts. We define five distinct *Nobecovirus* clades in global circulation across Asia and Africa and assess these new *Nobecoviruses* for their past and future capacity for recombination.

**Materials and Methods**

*Bat Sampling*

As part of a longterm study characterizing the seasonal dynamics of potentially zoonotic viruses in wild Madagascar fruit bats, monthly captures of Malagasy pteropodid bats were carried out at species-specific roost sites in the Districts of Moramanga and Manjakandriana, Madagascar between 2018 and 2019 (*P. rufus:* Ambakoana roost, -18.513 S, 48.167 E; *E. dupreanum*: AngavoBe cave, -18.944 S, 47.949 E; AngavoKely cave = -18.933 S, 47.758 E; *R. madagascariensis*:Maromizaha Cave,-18.9623 S, 48.4525 E). In brief, bats were captured in nets hung in the tree canopy (*P. rufus*) or over cave mouths (*E. dupreanum, R. madagascariensis)* at dusk (17:00-22:00) and dawn (03:00-07:00), removed from nets, and processed under manual restraint following methods that have been previously described (54,71,72). Briefly, all animals were identified to species, sex, and age class (juvenile vs. adult), and fecal, throat, and urine swabs were taken from each individual, collected into viral transport medium, and frozen on site in liquid nitrogen. Post-sampling, swabs were transported to -80\*C freezers for longterm storage in the Virology Unit at Institut Pasteur of Madagascar.

This study was carried out ﻿in strict accordance with research permits obtained from the Madagascar Ministry of Forest and the Environment (permit numbers 019/18, 170/18, 007/19) and under guidelines posted by the American Veterinary Medical Association. All field protocols employed were pre-approved by the UC Berkeley Animal Care and Use Committee (ACUC Protocol # AUP-2017-10-10393), and every effort was made to minimize discomfort to animals.

*RNA Extraction*

RNA was extracted from a randomly selected subset of fecal (285), throat (143), and urine (196) swab samples in the Virology Unit at Institut Pasteur of Madagascar, with each sample corresponding to a unique individual from the field dataset. Samples undergoing mNGS corresponded to individuals captured in Feb-Apr, Jul-Sep and December 2018 or in January 2019. Water controls were extracted in conjunction with samples on each unique extraction day. Extractions were conducted using the Zymo Quick DNA/RNA Microprep Plus kit (ZYMO, Irvine, CA, USA), according to the manufacturer’s instructions and including the step for DNAse digestion. Post-extraction, RNA quality was checked on a nanodrop to ensure that all samples demonstrated 260/280 ratios exceeding 2 and revealed quantifiable concentrations. Resulting extractions were stored in freezers at -80\*C, then transported on dry ice to the Chan Zuckerberg Biohub (San Francisco, CA, USA) for library preparation and metagenomic Next Generation Sequencing (mNGS).

*Library Preparation and mNGS*

A subset of four randomly selected samples from each of three bat species was selected for additional quantification using an Invitrogen Qubit 3.0 Fluorometer and the Qubit RNA HS Assay Kit (ThermoFisher Scientific, Carlsbad, CA, USA). After quantification, 5ul of of each RNA sample, plus water control, was diluted 5X and arrayed in 96-to-384 well plate format using a BRAVO Automated Liquid-Handling Platform and unique TruSeq Index PCR Primer barcodes (Illumina, San Diego, CA, USA). Samples were subsequently prepped into libraries using the NEBNext Directional RNA Library Prep Kit (Purified mRNA or rRNA Depleted RNA protocol; New England BioLabs, Beverly, MA, USA), following the manufacturer’s instructions and according to previously published modifications (XXX). Quality and quantity of resulting individual and pooled mNGS libraries were assessed via electrophoresis with the High Sensitivity NGS Fragment Analysis Kit on a Fragment Analyzer (Advanced Analytical Technologies, Inc), the High-Sensitivity DNA Kit on the Agilent Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA), and via real-time quantitative polymerase chain reaction (qPCR) with the KAPA Library Quantification Kit (Kapa Biosystems, Wilmington, MA, USA). Final library pools were spiked with a non-indexed PhiX control library (Illumina, San Diego, CA, USA). Pair-end sequencing (2 × 150 bp) was performed using an Illumina NovaSeq sequencing system (Illumina, San Diego, CA, USA). The pipeline used to separate the sequencing output into 150-base-pair pair-end read FASTQ files by library and to load files onto an Amazon Web Service (AWS) S3 bucket is available on GitHub at <https://github.com/czbiohub/utilities>.

*IDSeq*

Raw reads from Illumina sequencing were host-filtered, quality-filtered, and assembled on the IDseq (v3.2) platform, a cloud-based, open-source bioinformatics platform designed for microbe detection from metagenomic data (73), using a host background model compiled from publicly available full-length genomes for all three bat species. Samples were deemed “positive” for coronavirus infection if IDseq successfully assembled at least two nucleotide or protein-BLAST derived contigs with an average read depth >2 reads/nucleotide, mapping to any CoV reference accession number. To clarify that no positives were missed from IDseq, all non-host contigs assembled in IDseq underwent offline BLASTn and BLASTx (74) against a reference database constructed from all full-length reference sequences and protein reference sequences for *Alpha-* and *Betacoronavirus* available in NCBI Virus. Step-by-step instructions for our offline BLAST protocol can be accessed in our publically available GitHub repository at: https://github.com/brooklabteam/Mada-Bat-CoV/.

*Genome Annotation and BLAST*

Three full genome-length Nobecovirus contigs returned from IDseq (two from *R. madagascariensis* and one from *P. rufus*) were aligned with Nobecovirus homologs from NCBI (see ‘Phylogenetic Analysis’) and annotated in the program Geneious Prime (2020.0.5). We then used NCBI BLAST and BLASTx to query identity of our full length recovered genomes and their respective translated proteins to publicly available sequences in NCBI (74). We queried identity to reference sequences for four previously described *Nobecovirus* strains (accession numbers: MG762674 (HKU9), NC\_030886 (GCCDC1), MK211379 (GX2018), and NC\_048212 (*Eidolon helvum* sequences), as well as to the top BLAST hit overall. In instances where a putative gene recovered no hits to homologs via BLASTx, we instead queried the HHPred interactive server for protein and homology detection (75).

*Phylogenetic Analysis*

Contigs returned from IDseq were next combined with publicly available coronavirus sequences in NCBI to undertake phylogenetic analysis. We carried out three major phylogenetic analyses, building (a) a full-genome *Betacoronvirus* phylogeny, (b) a *Betacoronavirus* RdRp phylogeny corresponding to a conserved 259 bp fragment of the RNA-dependent RNA polymerase gene encapsulated in the CoV Orf1b, and (c) four amino acid phylogenies derived from translated nucleotides corresponding to the spike (S), envelope (E), matrix (M), and nucleocapsid (N) proteins of a subset of full length genomes. Detailed methods for the construction of each phylogeny are available at <https://github.com/brooklabteam/Mada-Bat-CoV/>.

Briefly, our full genome phylogeny was comprised of 122 unique NCBI records, corresponding to all available full genome sequences with bat hosts under NCBI taxon ids, *Betacoronavirus* (694002), unclassified *Betacoronavirus* (696098), *Betacoronavirus* sp. (1928434), unclassified Coronaviridae (1986197), or unclassified Coronavirinae (2664420) (107 records), in addition to all full genome *Betacoronavirus* (694002)**reference** sequences with a non-bat host (14 records), plus one *Gammacoronavirus* outgroup (accession number NC\_010800). The full genome phylogeny additionally included three full length Madagascar Nobecovirus sequences returned from IDseq (two from *R. madagascariensis* and one from *P. rufus*), which are described in this paper for the first time.

Our *Betacoronavirus* RdRp phylogeny consisted of an overlapping subset of a 259 bp RdRp fragment derived from *Betacoronviruses* previously described in Madagascar fruit bats (55) (7 records), in addition to the same RdRp fragment extracted from 17 near-full length *Nobecovirus* sequences, two RdRp *Nobecovirus* fragments, and 17 full length reference sequences for other *Betacoronavirus* subgenera available in NCBI Virus. This phylogeny also included seven Madagascar Nobecovirus sequences encompassing the RdRp fragment of interest, which were returned from the assembly in IDseq (four from *R. madagascariensis,* two from *P. rufus,* and one from *E. dupreanum*), in addition to the RdRp fragment of our *Gammacoronviru*s outgroup.

Lastly, our amino acid phylogenies consisted of S, E, M, and N gene extractions from the same representative set of near-full genome length sequences used in the RdRp analysis: the same 17 full-length *Betacoronavirus* reference sequences, 17 near full-length *Nobecovirus* sequences, and the one *Gammacoronavirus* outgroup, in addition to our three full genome Madagascar sequences derived from *R. madagascariensis* and *P. rufus*. Gene extractions were carried out using annotation tracks reported with each accession number in NCBI or, in cases where annotations were unavailable, genes were manually annotated and extracted in Geneious Prime based on alignment to homologs. After nucleotide extraction, genes were translated prior to alignment.

After compiling sequences for each disparate phylogenetic analysis, sequence subsets for the full-length, RdRp, and four amino acid phylogenies were aligned in MAFFT v.7 (76,77) using default parameter values. Alignments were checked manually for quality in Geneious Prime, and the RdRp aligment was trimmed to a 259 bp fragment conserved across all sequences in the subset. All sequence subsets and alignment files are available for public access in our Github repository: <https://github.com/brooklabteam/Mada-Bat-CoV/>.

After quality control, alignments were sent to Modeltest-NG (78) to assess the best fit nucleotide or amino acid substitution model appropriate for the data, then to RAxML-NG (79) to construct the corresponding maximum likelihood (ML) tree. Following best practices outlined in the RAxML-NG manual, twenty ML inferences were made on each original alignment and bootstrap replicate trees were inferred using Felsenstein’s method (80), with the MRE-based bootstopping test applied after every 50 replicates (81). Bootstrapping was terminated once diagnostic statistics dropped below the threshold value and support values were drawn on the best-scoring tree. Resulting phylogenies were visualized in R v.4.0.3 for MacIntosh, using the package ‘ggtree’ (82).

*Recombination Analysis*

Full length *Nobecovirus* sequences derived from IDseq were analyzed for any signature of past recombination. All three full length Madagascar *Nobecovirus* genomes were aligned with full genome sequences corresponding to two disparate *Nobecovirus* lineages: the HKU9 lineage (EF065514-EF065516, HM211098-HM211100, MG693170, NC\_009021, MG762674) and the *Eidolon helvum* Africa lineage (MG693169, MG693171-MG693172, NC\_048212). Alignment was conducted in MAFFT v.7 (76,77) using default parameter values. *Nobecovirus* sequences corresponding to the GCCDC1 (21,60) and GX2018/BatCoV92 (103,132) genotypes were omitted from recombination analyses because inserted genes and/or genetic material upstream from the nucleocapsid in the corresponding genomes interfered with the alignment.

After alignment, genomes were analyzed for recombination in the program SimPlot (v.3.5.1). Similarity plots, which compute identity across all genomic positions in an alignment, were generated using the *P. rufus* and, subsequently, the *R. madagascariensis* genomes as query sequences, the HKU9 and *Eidolon helvum* African *Nobecovirus* clades as references, and the corresponding Madagascar sequence as the alternative. Bootscan analyses were conducted on the same alignment, using the same query and reference inputs. Both Similarity and Bootscan analyses were carried out using a window size of 200bp and a step size of 20bp.

*Nucleotide Sequence Accession Numbers*

All three annotated full-length genome sequences (two from *R. madagascariensis,* one from *P. rufus*), plus four additional RdRp gene fragment sequences (two from *R. madagascariensis,* one from *P. rufus*, and one from *E. dupreanum*) were submitted to NCBI and assigned accession numbers XXXXX-XXXX (pending).

**Results**

*Coronavirus Prevalence in Field Samples*

IDseq assembly generated at least two nucleotide or protein-BLAST derived contigs with an average read depth > 2 reads/nucleotide, mapping to any CoV reference accession number, in 28/285 (9.82%) of fecal samples and in 2/196 (1.00%) of urine samples. Given low prevalence, the latter urine positives likely signified field contamination with fecal excrement upon urine swab collection, as bats often excrete both substances simultaneously under manual restraint. None of the 143 throat swabs assayed demonstrated evidence of CoV infection.

Prevalence in feces varied slightly across species, with 4/44 (9.09%) of *P. rufus* samples*,* 16/145 (11.03%) of *E. dupreanum* samples, and 8/96 (8.33%) of *R. madagascariensis* samples testing positive for CoV infection. Juveniles demonstrated higher CoV prevalence than adults for *P. rufus* and *E. dupreanum* but not for *R. madagascariensis*. Juvenile vs. adult prevalence was 3/15 (20%) vs. 1/29 (3.45%) for *P. rufus,*  5/13 (38.46%) vs. 11/132 (8.33%) for *E. dupreanum*, and 0/13 (0%) vs. 8/83 (9.64%) for *R. madagascaiensis* (**Figure 1**)*.*

*Genome Annotation and BLAST*

Three full genome or near-full genome length contigs were recovered from IDseq for *Nobecoviruses* derived from *R. madagascariensis* (two genomes: 28,980 and 28,926 bps in length) and *P. rufus* (one genome: 29,122 bps in length). In all three genomes, we successfully annotated ORF1ab (including RdRp) and structural proteins S (spike), E (envelope), M (matrix), and N (nucleocapsid), in addition to accessory genes NS3, NS7a, and NS7b (**Figure 2**). Genomes derived from *R. madagascariensis* appeared slightly more complex than those derived from *P. rufus,* allowing for annotation of one additional accessory gene, NS7c, which has been described previously in recombinant *Nobecovirus* sequences of the GCCDC1 lineage (37,105).

In addition to major proteins, we successfully identified several 6 bp motifs corresponding to the 5’-ACGAAC-3’ core Transcription Regulatory Sequence (TRS) common to many *Betacoronaviruses,* including SARS-CoV and previously described in *Nobecoviruses* of the GCCDC1 and GX2018/BatCoV92 lineages (132–134). For most genes, these TRS elements were located a short distance upstream from the corresponding gene (**Table 1**). Elements identified in the two *R. madagascariensis* genomes were largely comparable, suggesting that these two sequences could represent slight variations in the same virus lineage—though the intergenic region for one putative *R. madagascariensis* TRS located upstream from the N gene was substantially longer than the other, thus highlighting the dynamic nature of the CoV genome in this region. This region corresponds to the site of the previously described orthoreovirus gene insertion in *Nobecovirus* genotype GCCDC1 (133). Some putative TRS elements showed variation from the 5’-ACGAAC-3’ core motif (including some which recapitulated the 5’-AAGAA-3’ motif common to SARS-CoV-2 (135)). TRS variations may be indicative of variation in gene expression across individual bats and/or species.

BLAST analysis of the full genome indicated that the *P.rufus* *Nobecovirus* sequence is highly divergent, demonstrating 72-73% identity to all previously described *Nobecovirus* clades, with the top blast association to HKU9 (**Table S1**). By contrast, *Nobecovirus* genomes derived from *R. madagascariensis* demonstrated high identity (~95%) to *Eidolon helvum Nobecovirus* lineages circulating in Africa. BLASTx analysis of individual genes from viruses derived from both Madagascar species demonstrated the highest identity with previously described *Nobebecovirus* sequences in the Orf1b region (which includes RdRp) for both *P. rufus* and *R. madagascariensis* viruses (70.08% identity for *P. rufus* to HKU9and 99% identity for *R. madagascariensis* to *E. helvum*). By contrast, both *P. rufus* and *R. madagascariensis* genomes demonstrated substantial divergence from all known homologs in the S and NS7 regions of their genomes, showing only 45.59-63.64% identity to the closest hits across these regions. The *P. rufus Nobecovirus* was similarly divergent in the N gene, though *R. madagascariensis Nobecoviruses* demonstrated high (86-87%) identity to *E. helvum* genotypes in this region.

*Phylogenetic Analysis*

Phylogenetic analysis of full length *Betacoronavirus* genomes confirmed that both *P. rufus* and *R. madagascariensis* genomes cluster in the *Nobecovirus* subgenus of the *Betacoronaviruses*, with the divergent *P. rufus* forming its own distinct clade and both *R. madagascariensis* genomes grouping with the previously described *Eidolon helvum* reference sequence from Cameroon (102) (**Figure 3A**). We observed distinct groupings of five main *Nobecovirus* lineages in our phylogeny: (a) the largely Asian-derived HKU9 sequences, (b) the African *E. helvum-*derivedsequences (now including new *R. madagascariensis* genomes), (c) the orthereovirus recombinant GCCDC1 genomes, (d) the GX2018/BatCoV92 genomes described from China and Singapore, and (e) the divergent *P. rufus* genome contributed here from Madagascar. Intriguingly, the *P. rufus* genome groups ancestral to all other *Nobecoviruses,* followed by the *E. helvum/R. madagascariensis* African lineage, with the Asian genotypes forming three distinct (and more recent) clades corresponding to genotypes HKU9, GCCDC1, and Gx2018/BatCoV92.Further phylogenetic analysis of a 259bp fragment of the RdRp gene reconfirmed these groupings and suggested the presence of at least two distinct genetic variants within the *P. rufus* lineage (**Figure 3B**). One RdRp fragment derived from feces of the third Malagasy fruit bat, *Eidolon dupreanum,* grouped within the *E. helvum/R. madagascariensis* African *Nobecovirus* lineage*,* consistent with previous reporting (100). Characterization of the full length genome of this virus will be needed to clarify whether it represents a genetic variant of or a distinct genotype from the *R. madagascariensis* virus. Phylogenetic analysis of the RdRp fragment allowed for inclusion of one partial *Nobecovirus* sequence derived from *E. helvum* bats in Kenya (HQ728482), which also grouped within the *E. helvum/R. madagascariensis* African clade, confirming the distribution of this genotype across West and East Africa and into the Indian Ocean*.* Notably, one partial Cameroonian *E. helvum* sequence (MG693170) clustered with HKU9 sequences from Asia, rather than within the *E. helvum/R. madagascariensis* African clade. These findings suggest that both “African” and “Asian” *Nobecovirus* lineages are likely broadly geographically distributed.

Amino acid phylogenies computed from translated protein alignments of the S, E, M, and N *Betacoronavirus* structural genes (**Figure 4 A-D**, respectively) further confirmed evolutionary relationships suggested in Figure 3. S, M, and N gene phylogenies demonstrated distinct groupings of five main *Nobecovirus* lineages outlined above, while in the E gene phylogeny, the *P. rufus* sequence grouped adjacent to the single Cameroonian-derived *E. helvum* sequence within the HKU9 clade.

*Recombination Analysis*

SimPlot analysis confirmed the evolutionary distinctiveness of the *P. rufus Nobecovirus* genome, which showed <50% identity to HKU9, *E. helvum,* and *R. madagascariensis* genotypes across the majority of its genome length (**Figure 5A**). Consistent with BLAST results, the *P. rufus Nobecovirus* genome demonstrated the highest similarity to previously described sequences in the Orf1b region, which includes RdRp. The *R. madagascariensis Nobecoviruses*, by contrast, showed >90% similarity to the *E. helvum* African lineage throughout Orf1ab, but both *P. rufus* and *R. madagascariensis* sequences diverged from all other reference genomes in the first half of the spike protein, which corresponds to the S1 subunit and includes the receptor binding domain that mediates viral entry into host cells (136). Further divergence for both *P. rufus* and *R. madagascariensis Nobecoviruses* was observed in the N structural protein and in the NS7 accessory genes. Bootscan analysis further confirmed these findings, showing that the *P. rufus Nobecovirus* clusters with HKU9 lineages across Orf1ab, NS3, E, and M genes but demonstrates evidence of recombination with *E. helvum/R. madagascariensis* African lineages in the S (particularly S1), N, and NS7 genes (**Figure 5B**). Similarly, bootscanning demonstrated that *R. madagascariensis Nobecoviruses* group with the *E. helvum* lineage across Orf1ab, NS3, E, and M but show evidence of recombination with HKU9 and *P. rufus Nobecovirus* in S (again, particularly S1), N, and NS7 genes (**Figure 5B**), thus highlighting the dynamic nature of these regions of the Nobecovirus genome.

**Discussion (IN PROGRESS)**

* We define Nobecov clades!
* CoV tropism appears largely gastrointestinal in bats, as compared with respiratory in humans
* staggered seasonality, with Rousettus birthing latest… Rousettus juveniles not infected… is this because they were sampled too early when still maternally immune and later-stage juveniles are missed because they resemble adults? (most R.mad juveniles were sampled in Feb, at only ~1 month old)
* sampling CoVs in Australia (other pteropus bats) would be helpful to describe the geographic breadth of NobeCovs
* S1 subunit divergence in spike is notable
* Also notable that there is a signature of recombination in NS7…what is its function? Immune evasion?
* Also notable that R.mad and P. ruf show recombination in the same regions of the genome. These are dynamic regions

Organize as:

1. Two novel Nobecos, cluster with Asian clades (Pteropus) and African (Rousettus) but evidence of recombination in S, N, NS7 genes
2. No evidence of orthereovirus insertion, suggests this strain may be limited to Asia. In fact, we can define four clades of Nobecoviruses broadly: HKU9, Eidolon helvum, GCCD1, and BatCoV92/GX2018.
3. Following on above, P. rufus does have extra genetic material between M and N, as does BatCoV92/GX2018, suggesting a dynamic region of the genome that could be a site for future recombination or acqusistion of new genes
4. Serious concern would be acquisition of S sequences enabling human cell entry. No known Sarbecoviruses on the island that could enable this but there are Merbecobivurses and M. jugalaris coroosts with Rousettus
5. Probably bigger concern is spillback and additional genetic material for SARS-CoV-2 which is widespread in Mada
6. All the seasonality stuff and importance of longitudinal studies
7. A plug for the importance of full genomes – only a handful of Nobeco genomes out there

We have described three novel *nobecovirus* sequences, most notably from *R. madagascariensis*. The average prevalence of 10% is comparable to sample efforts in other countries, indicating that there is an endemic level of coronaviruses circulating throughout Madagascar(73,74). A previous coronavirus sampling study of Madagascar fruit bats found viruses in *P. rufus* and *E. dupreanum*, but not *R. madagascariensis*, although they only detected one virus in *E. dupreanum*(74). Most of their sampling was also within a one year span, and mostly restricted to one region, which could explain the skewing of positive samples toward one bat species, but still resulted in an overall prevalence of 4.5%(74). Another study of coronavirus sampling in the West Indian Ocean provided more information about prevalence in Madagascar (around 5%) with a larger sample set that is more ubiquitously spread about the island, but also showed that the islands sampled have similar coronavirus prevalence to that of Africa28. Based on full genome phylogeny our novel viruses cluster closely with Asian *Rousettus* clades in general, both *R. madagascariensis* genomes clustered most closely to an *E. helvum* genome from Cameroon (Figure 3). Similar results were found looking at phylogeny of only the RdRp, further providing evidence that *R. madagascariensis* isolated are closely related to *Eidolon* viruses (Figure 3). This provides support for the evolutionary path these viruses followed, and most likely resulted from recombination events as demonstrated by evidence for this in the S, N, and NS7 genes, whereas nonstructural regions such as ORF1ab remained similar (Figure 4).

It has been hypothesized that some of the waves of the SARS-CoV-2 pandemic may have in part arisen through recombination events of North American and European strains(49).

Recombination among coronaviruses is documented in bats and is being observed more frequently, recombination events could be source of SARS-CoV-2, whereas SARS and MERS-CoV were thought to have infected humans through a direct zoonotic or through an intermediate zoonotic host(57–62). Roosting bat species are an obvious source of recombination between coronaviruses of one species but also co-roosting populations as well; in one mine shaft in China that hosted multiple species of bats displayed evidence of coinfection with coronaviruses along with novel *Betacoronaviruses* and a new *sarbecovirus* (63). Madagascar endemic species of bats can co-roost in the same habitats; *R. madagascariensis* and *E. dupreanum* roost in caves, whereas P. rufus roosts in trees(83). Recombination also has recently been show to occur with a p10 gene of bat orthoreovirus of *Pteropus* origin into a *Betacoronavirus* of the GCCDC1 lineage between N and NS7, showing similarity to Rousettus bat coronavirus HKU9(64). In previous Madagascar bat sampling and within our sample set, this orthoreovirus insertion was not detected, indicating that while belonging to the Asian lineage of coronaviruses may also be restricted geographically to Asia itself(17,35). However, there has been one report of an orthoreovirus inserstion in the same parent virus in the same location of the p10 orthoreovirus segment outside of mainland China, in captive Singapore *Eonycteris* bats(65). There are four known clades of *Nobecovirus* based on our phylogeny: *Rousettus* bat coronavirus HKU9, *Eidolon helvum* clade, *Rousettus* bat coronavirus GCCDC1, and BatCoV92/GX2018.

Genetically, *Nobecovirus* clade BatCoV92/GX2018 has extra genetic elements between the M and N genes, as also seen in our *P. rufus* genome(66). In fact, there is a transcription regulatory sequence (TRS), a start codon, and a stop codon. The TRS plays a role in the replication of the coronavirus genome, indicating where RdRp should switch templates, lending to the idea that TRS are a potential site of recombination in coronaviruses as once the template is switched, it could link to another genome(67–70). The TRS motif is also well conserved in *Betacoronaviruses* 5’-ACGAAC-3’, which we found in the majority of our TRS locations, with other *Betacoronavirus* TRS motifs found in some of the accessory proteins, such as NS7b and NS7c, as indicated in table 1. This ORFx region between M and N in our *P. rufus* genome may be susceptible to recombination, as it has a TRS region, and through BLAST results does not indicate a protein that is known to be part of the normal coronavirus genome. While not in the same location as the p10 orthoreovirus insertion, this ORFx region is close in genome location, and could serve as a potential site. *Pteropine* orthoreovirus is found in Madagascar fruit bats (unpublished data). Other RNA viruses circulate in Madagascar as well as these *Nobecoviruses*, such as Henipaviruses and paramyxoviruses, and paramyxoviruses in particular are susceptible to recombination(53,71–73).

An important recombination event would be one that gave a bat coronavirus a S sequence that would allow for human cell entry, as is predicted to have happened in the emergence of SARS-CoV-2(25,57,59,74). One study found a coronaviruses in Africa that appears to be an intermediate step between SARS-CoV-1 and SARS-CoV-2 in terms of similarity in the receptor binding domain, but without the ability to bind ACE2(99). ACE2 usage is well described in many coronaviruses from Asia, but more focus should be on bridging the gap in this knowledge from other countries(78,99). As of now there are no known circulating *Sarbecoviruses* in Madagascar fruit bats, which would be the greatest concern, but there are circulating *Hibecoviruses* in *Hipposideridae* and *Merbecoviruses* in *Vespertilionidae* which may co-roost with *Nobecovirus* associated species(97). The SARS-CoV-2 pandemic has taken hold in Madagascar, and while not in a surge at time of writing, remains a country with limited access to vaccination which allows the virus to continue spreading throughout a susceptible population(75). While *Sarbecoviruses* are not currently found in Madagascar bats, the possibility of reverse spillover from the human population into the bats should be considered in this isolated island population where there may be an evolutionary pressure to jump to an ubiquitous and potentially susceptible host. Analysis on reverse spillover has been done to assess the risk to North American bats, but not in other at-risk countries that have more contact with bats in daily life such as Africa, Madagascar, and Asia(50). In fact, mink-adapted SARS-CoV-2 in Finland already jumped back into farmers that were in close contact with the animals(76).

Seasonality modeling of coronavirus prevalence revealed little data to correlate infection data to bat breeding seasons and annual birth pulses, so more data is needed to correlate the time of year the sample was collected to food availability, depending on the species’ diet. Pathogen spillover from bats is also dictated by ecological factors such as seasonality, waning immunity, and other stressors such as nutrition access and breeding seasons(73,100). In our study, the highest prevalence of coronaviruses occurred between March-April, leading up to the breeding season for the three bat species. Stress in these bat species my also dictate coronavirus success in these hosts, as stress can dampen the immune response46. Multi-year longitudinal studies will be necessary to untangle these interactions. Data on human risk from these coronaviruses is lacking. Bats come into contact with humans on Madagascar through habitat destruction along with through hunters, several bat species are consumed(73,101–103). Close contact with roosting habitats such as caves not only puts a human at risk of direct bat contact, but also with guano. In addition to longitudinal sampling of bats, it would be beneficial to supplement this data with antibody studies from local human populations such as hunters to assess zoonotic risk, with a particular focus on coronaviruses along with other pathogens of interest such as henipaviruses that are shown to replicate in these species discussed(83). With how ubiquitous bats are, it is important to recognize the risk while also understanding that they are important members of many ecosystems, and protection from habitat loss and encroachment would go a long way in preventing unnecessary human/bat interactions. With little information on *Nobecoviruses* and few full isolated genomes, surveillance work in susceptible hosts worldwide is important to try and predict the next potential spillover and understand ecological pathogen dynamics as they naturally exist.

**Conflict of Interest:**

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

**Author Contributions:**

**Funding:**

**Acknowledgements:**

**Data Availability Statement:**

**References**

1. Ravelomanantsoa NAF, Guth S, Andrianiaina A, Andry S, Gentles A, Ranaivoson HC, et al. The zoonotic potential of bat-borne coronaviruses. Emerging Topics in Life Sciences. 2020;In Press.

2. Banerjee A, Kulcsar K, Misra V, Frieman M, Mossman K. Bats and coronaviruses. Viruses. 2019;11(1):7–9.

3. Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020;

4. Hu B, Ge X, Wang LF, Shi Z. Bat origin of human coronaviruses. Virology Journal. 2015;12(1):1–10.

5. Wille M, Holmes EC. Wild birds as reservoirs for diverse and abundant gamma- And deltacoronaviruses. FEMS Microbiology Reviews. 2020;44(5):631–44.

6. Drexler JF, Gloza-Rausch F, Glende J, Corman VM, Muth D, Goettsche M, et al. Genomic characterization of severe acute respiratory syndrome-related coronavirus in European bats and classification of coronaviruses based on partial RNA-dependent RNA polymerase gene sequences. Journal of virology. 2010/08/04. 2010 Nov;84(21):11336–49.

7. Hu B, Zeng L-P, Yang X-L, Ge X-Y, Zhang W, Li B, et al. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. PLoS pathogens. 2017 Nov 30;13(11):e1006698–e1006698.

8. Anthony SJ, Gilardi K, Menachery VD, Goldstein T, Ssebide B, Mbabazi R, et al. Further evidence for bats as the evolutionary source of middle east respiratory syndrome coronavirus. mBio. 2017;8(2):1–13.

9. Woo PC, Lau SK, Li KS, Tsang AK, Yuen K-Y. Genetic relatedness of the novel human group C betacoronavirus to Tylonycteris bat coronavirus HKU4 and Pipistrellus bat coronavirus HKU5. Emerging microbes & infections. 2012/11/07. 2012 Nov;1(11):e35–e35.

10. Corman VM, Ithete NL, Richards LR, Schoeman MC, Preiser W, Drosten C, et al. Rooting the phylogenetic tree of middle East respiratory syndrome coronavirus by characterization of a conspecific virus from an African bat. Journal of virology. 2014/07/16. 2014 Oct;88(19):11297–303.

11. Razanajatovo NH, Nomenjanahary LA, Wilkinson DA, Razafimanahaka JH, Goodman SM, Jenkins RK, et al. Detection of new genetic variants of Betacoronaviruses in endemic frugivorous bats of Madagascar. Virology Journal. 2015;12(42).

12. P LSK, S PRW, L WBH, Ming W, Yi H, Huifang X, et al. Coexistence of different genotypes in the same bat and serological characterization of Rousettus bat coronavirus HKU9 belonging to a novel Betacoronavirus subgroup. Journal of Virology. 2010 Nov 1;84(21):11385–94.

13. Frutos R, Serra-Cobo J, Pinault L, Lopez Roig M, Devaux CA. Emergence of bat-related Betacoronaviruses: Hazard and risks. Frontiers in Microbiology. 2021;12:437.

14. Chen S-C, Olsthoorn RCL, Yu C-H. Structural phylogenetic analysis reveals lineage-specific RNA repetitive structural motifs in all coronaviruses and associated variations in SARS-CoV-2. Virus Evolution. 2021 Jan 20;7(1).

15. Zhou H, Chen X, Hu T, Li J, Song H, Liu Y, et al. A novel bat coronavirus reveals natural insertions at the S1/S2 cleavage site of the Spike protein and a possible recombinant origin of HCoV-19. bioRxiv. 2020 Jan 1;2020.03.02.974139.

16. Forni D, Cagliani R, Sironi M. Recombination and positive selection differentially shaped the diversity of Betacoronavirus subgenera. Viruses. 2020 Nov 16;12(11):1313.

17. Llanes A, Restrepo CM, Caballero Z, Rajeev S, Kennedy MA, Lleonart R. Betacoronavirus genomes: How genomic information has been used to deal with past outbreaks and the COVID-19 pandemic. International journal of molecular sciences. 2020 Jun 26;21(12):4546.

18. Li W, Shi Z, Yu M, Ren W, Smith C, Epstein JH, et al. Bats are natural reservoirs of SARS-like coronaviruses. Science. 2005 Oct 28;310(5748):676.

19. Lam TT-Y, Jia N, Zhang Y-W, Shum MH-H, Jiang J-F, Zhu H-C, et al. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. Nature. 2020;583(7815):282–5.

20. Hul V, Delaune D, Karlsson EA, Hassanin A, Tey PO, Baidaliuk A, et al. A novel SARS-CoV-2 related coronavirus in bats from Cambodia. bioRxiv. 2021 Jan 1;2021.01.26.428212.

21. Paskey AC, Ng JHJ, Rice GK, Chia WN, Philipson CW, Foo RJH, et al. Detection of recombinant Rousettus bat coronavirus GCCDC1 in lesser dawn bats (Eonycteris spelaea) in Singapore. Viruses. 2020 May 14;12(5):539.

22. Valitutto MT, Aung O, Tun KYN, Vodzak ME, Zimmerman D, Yu JH, et al. Detection of novel coronaviruses in bats in Myanmar. PLOS ONE. 2020 Apr 9;15(4):e0230802-.

23. Lau SKP, Li KSM, Huang Y, Shek C-T, Tse H, Wang M, et al. Ecoepidemiology and complete genome comparison of different strains of severe acute respiratory syndrome-related Rhinolophus bat coronavirus in China reveal bats as a reservoir for acute, self-limiting infection that allows recombination events. Journal of virology. 2010/01/13. 2010 Mar;84(6):2808–19.

24. Latinne A, Hu B, Olival KJ, Zhu G, Zhang L, Li H, et al. Origin and cross-species transmission of bat coronaviruses in China. Nature Communications. 2020;11(1):4235.

25. Wacharapluesadee S, Duengkae P, Rodpan A, Kaewpom T, Maneeorn P, Kanchanasaka B, et al. Diversity of coronavirus in bats from Eastern Thailand. Virology Journal. 2015;12(1):57.

26. Ying T, Mang S, Christina C, Krista Q, Jing Z, Wanda M, et al. Surveillance of bat coronaviruses in Kenya identifies relatives of human coronaviruses NL63 and 229E and their recombination history. Journal of Virology. 2021 Aug 4;91(5):e01953-16.

27. Montecino-Latorre D, Goldstein T, Gilardi K, Wolking D, Van Wormer E, Kazwala R, et al. Reproduction of East-African bats may guide risk mitigation for coronavirus spillover. One Health Outlook. 2020;2(1):2.

28. Tong S, Conrardy C, Ruone S, Kuzmin I V, Guo X, Tao Y, et al. Detection of novel SARS-like and other coronaviruses in bats from Kenya. Emerging infectious diseases. 2009 Mar;15(3):482–5.

29. Joffrin L, Goodman SM, Wilkinson DA, Ramasindrazana B, Lagadec E, Gomard Y, et al. Bat coronavirus phylogeography in the Western Indian Ocean. Scientific Reports. 2020;10(1):6873.

30. Anthony SJ, Ojeda-Flores R, Rico-Chávez O, Navarrete-Macias I, Zambrana-Torrelio CM, Rostal MK, et al. Coronaviruses in bats from Mexico. Journal of General Virology. 2013;94(PART 5):1028–38.

31. Su S, Wong G, Shi W, Liu J, Lai ACK, Zhou J, et al. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. Trends in Microbiology. 2016;24(6):490–502.

32. Woolhouse MEJ, Haydon DT, Antia R. Emerging pathogens: the epidemiology and evolution of species jumps. Trends in Ecology and Evolution. 2005 May;20(5):238–44.

33. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020;

34. Li W, Moore MJ, Vasilieva N, Sui J, Wong SK, Berne MA, et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature. 2003;147(2):120–1.

35. Raj VS, Mou H, Smits SL, Dekkers DHW, Müller MA, Dijkman R, et al. Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. Nature. 2013;495(7440):251–4.

36. Tao Y, Tong S. Complete genome sequence of a Severe Acute Respiratory Syndrome-related Coronavirus from Kenyan bats. Microbiology Resource Announcements. 2019;8(28):e00548-19.

37. Wells HL, Letko M, Lasso G, Ssebide B, Nziza J, Byarugaba DK, et al. The evolutionary history of ACE2 usage within the coronavirus subgenus sarbecovirus. Virus Evolution. 2021 Jan 20;7(1).

38. Lau SKP, Feng Y, Chen H, Luk HKH, Yang W-H, Li KSM, et al. Severe Acute Respiratory Syndrome (SARS) coronavirus ORF8 protein is acquired from SARS-related coronavirus from greater horseshoe bats through recombination. Journal of Virology. 2015;89(20):10532–47.

39. Vijgen L, Keyaerts E, Moes E, Thoelen I, Wollants E, Lemey P, et al. Complete genomic sequence of human coronavirus OC43: Molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event. Journal of Virology. 2005;79(3):1595–604.

40. Corman VM, Eckerle I, Memish ZA, Liljander AM, Dijkman R, Jonsdottir H, et al. Link of a ubiquitous human coronavirus to dromedary camels. Proceedings of the National Academy of Sciences. 2016;201604472.

41. Corman VM, Baldwin HJ, Tateno AF, Zerbinati RM, Annan A, Owusu M, et al. Evidence for an ancestral association of human coronavirus 229E with bats. Journal of Virology. 2015;89(23):11858–70.

42. Graham RL, Baric RS. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. Journal of virology. 2009/11/11. 2010 Apr;84(7):3134–46.

43. Ravelomanantsoa NAF, Guth S, Andrianiaina A, Andry S, Gentles A, Ranaivoson HC, et al. The zoonotic potential of bat-borne coronaviruses. Emerging Topics in Life Sciences. 2020 Dec 11;4(4).

44. Ogando NS, Ferron F, Decroly E, Canard B, Posthuma CC, Snijder EJ. The curious case of the Nidovirus exoribonuclease: Its role in RNA synthesis and replication fidelity. Frontiers in Microbiology. 2019;10:1813.

45. Nga PT, de Parquet MC, Lauber C, Parida M, Nabeshima T, Yu F, et al. Discovery of the first insect nidovirus, a missing evolutionary link in the emergence of the largest RNA virus genomes. PLoS Pathogens. 2011;7(9):e1002215.

46. Gorbalenya AE, Enjuanes L, Ziebuhr J, Snijder EJ. Nidovirales: Evolving the largest RNA virus genome. Virus Research. 2006;117(1):17–37.

47. Smith EC, Blanc H, Vignuzzi M, Denison MR. Coronaviruses lacking exoribonuclease activity are susceptible to lethal mutagenesis: Evidence for proofreading and potential therapeutics. PLoS Pathogens. 2013;9(8).

48. Lai MMC. RNA recombination in animal and plant viruses. Microbiological Reviews. 1992;56(1):61–79.

49. Masters J, Wit M De, Asher RJ. Reconciling the origins of Africa, India and Madagascar with vertebrate dispersal scenarios. Folia Primatologica. 2006;3200:399–418.

50. Species IUCN Red List Threat. IUCN 2018. Version 2018-2.

51. Almeida FC, Giannini NP, Simmons NB, Helgen KM. Each flying fox on its own branch: a phylogenetic tree for Pteropus and related genera (Chiroptera: Pteropodidae). Molecular Phylogenetics and Evolution. 2014 Mar;(March).

52. Shi JJ, Chan LM, Peel AJ, Lai R, Yoder AD, Goodman SM. A deep divergence time between sister species of *Eidolon* (Pteropodidae) with evidence for widespread panmixia. Acta Chiropterologica. 2014 Dec;16(2):279–92.

53. Goodman SM, Chan L, Nowak M, Yoder AD. Phylogeny and biogeography of western Indian Ocean *Rousettus* (Chiroptera : Pteropodidae). Journal of Mammalogy. 2010;91(3):593–606.

54. Brook CE, Ranaivoson HC, Broder CC, Cunningham AA, Héraud J-M, Peel AJ, et al. Disentangling serology to elucidate henipa- and filovirus transmission in Madagascar fruit bats. Journal of Animal Ecology. 2019 Jul 1;88(7):1001–16.

55. Razanajatovo NH, Nomenjanahary LA, Wilkinson DA, Razafimanahaka JH, Goodman SM, Jenkins RK, et al. Detection of new genetic variants of Betacoronaviruses in endemic frugivorous bats of Madagascar. Virology Journal. 2015;12(1):42.

56. Reynes J-M, Andriamandimby SF, Razafitrimo GM, Razainirina J, Jeanmaire EM, Bourhy H, et al. Laboratory surveillance of rabies in humans, domestic animals, and bats in Madagascar from 2005 to 2010. Advances in preventive medicine. 2011 Jan;2011:727821.

57. Yinda CK, Ghogomu SM, Conceição-Neto N, Beller L, Deboutte W, Vanhulle E, et al. Cameroonian fruit bats harbor divergent viruses, including rotavirus H, bastroviruses, and picobirnaviruses using an alternative genetic code. Virus Evolution. 2018;4(1):1–15.

58. Lim XF, Lee CB, Pascoe SM, How CB, Chan S, Tan JH, et al. Detection and characterization of a novel bat-borne coronavirus in Singapore using multiple molecular approaches. Journal of General Virology. 2019;100(10):1363–74.

59. Han Y, Du J, Su H, Zhang J, Zhu G, Zhang S, et al. Identification of diverse bat Alphacoronaviruses and Betacoronaviruses in china provides new insights into the evolution and origin of Coronavirus-related diseases. Frontiers in Microbiology. 2019;10(AUG).

60. Huang C, Liu WJ, Xu W, Jin T, Zhao Y, Song J, et al. A bat-derived putative cross-family recombinant coronavirus with a reovirus gene. PLoS Pathogens. 2016;12(9):1–25.

61. Obameso JO, Li H, Jia H, Han M, Zhu S, Huang C, et al. The persistent prevalence and evolution of cross-family recombinant coronavirus GCCDC1 among a bat population : a two-year follow-up. 2017;60(12):1357–63.

62. Goodman SM. Les chauves-souris de Madagascar [in French]. Antananarivo, Madagascar: Association Vahatra; 2011.

63. Jenkins RKB, Racey PA. Bats as bushmeat in Madagascar. Madagascar Conservation and Development. 2008;3(1):22–30.

64. Golden CD, Bonds MH, Brashares JS, Rodolph Rasolofoniaina BJ, Kremen C. Economic valuation of subsistence harvest of wildlife in Madagascar. Conservation Biology. 2014 Jan 9;1–10.

65. Randrianandrianina F, Andriafidison D, Amyot F, Ramilijaona O, Ratrimomanarivo F, Racey PA, et al. Habitat use and conservation of bats in rainforest and adjacent human-modified habitats in eastern Madagascar. Acta Chiropterologica. 2006;8(2):429–37.

66. Cardiff SG, Ratrimomanarivo FH, Goodman SM. The Effect of Tourist Visits on the Behavior of *Rousettus madagascariensis* (Chiroptera: Pteropodidae) in the Caves of Ankarana, Northern Madagascar. Acta Chiropterologica. 2012 Dec;14(2):479–90.

67. Razanajatovo NH, Richard V, Hoffmann J, Reynes J, Razafitrimo M, Randremanana RV, et al. Viral etiology of Influenza-like illnesses in Antananarivo, Madagascar, July 2008 to June 2009. 2011;6(3).

68. Razanajatovo NH, Guillebaud J, Harimanana A, Rajatonirina S, Ratsima EH, Andrianirina ZZ, et al. Epidemiology of severe acute respiratory infections from hospital-based surveillance in Madagascar, November 2010 to July 2013. PLoS ONE. 2018;(July 2013):1–17.

69. Randremanana R, Andriamandimby S, Rakotondramanga JM, Razanajatovo N, Mangahasimbola R, Randriambolamanantsoa T, et al. The COVID-19 Epidemic in Madagascar: clinical description and laboratory results of the first wave, March-September 2020. Influenza and Other Respiratory Viruses. 2021;00:1–12.

70. Olival KJ, Cryan PM, Amman BR, Baric RS, Blehert DS, Brook CE, et al. Possibility for reverse zoonotic transmission of SARS-CoV-2 to free-ranging wildlife: A case study of bats. PLOS Pathogens. 2020 Sep 3;16(9):e1008758-.

71. Ranaivoson HC, Héraud J-M, Goethert HK, Telford SR, Rabetafika L, Brook CE. Babesial infection in the Madagascan flying fox, Pteropus rufus É. Geoffroy, 1803. Parasites & Vectors. 2019;12(1):51.

72. Brook CE, Bai Y, Dobson AP, Osikowicz LM, Ranaivoson HC, Zhu Q, et al. Bartonella spp. in Fruit Bats and Blood-Feeding Ectoparasites in Madagascar. PLOS Neglected Tropical Diseases. 2015 Feb 23;9(2):e0003532-.

73. Kalantar KL, Carvalho T, De Bourcy CFA, Dimitrov B, Dingle G, Egger R, et al. IDseq-An open source cloud-based pipeline and analysis service for metagenomic pathogen detection and monitoring. GigaScience. 2021;9(10):1–14.

74. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of Molecular Biology. 1990;215(3):403–10.

75. Söding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Research. 2005;33(SUPPL. 2):244–8.

76. Kuraku S, Zmasek CM, Nishimura O, Katoh K. aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. Nucleic acids research. 2013;41(Web Server issue):22–8.

77. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. Briefings in Bioinformatics. 2018;20(4):1160–6.

78. Darriba Di, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. ModelTest-NG: A new and scalable tool for the selection of DNA and protein evolutionary models. Molecular Biology and Evolution. 2020;37(1):291–4.

79. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics. 2019;35(21):4453–5.

80. Felsenstein J. Confidence limits on phylogenies: An approach using the bootstrap. Evolution. 1985;39(4):783–91.

81. Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A. How many bootstrap replicates are necessary? Journal of Computational Biology. 2010;17(3):337–54.

82. Yu G, Smith DK, Zhu H, Guan Y, Lam TTY. Ggtree: an R Package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution. 2017;8(1):28–36.

83. Han Y, Du J, Su H, Zhang J, Zhu G, Zhang S, et al. Identification of diverse bat Alphacoronaviruses and Betacoronaviruses in China provides new insights into the evolution and origin of coronavirus-related diseases. Frontiers in Microbiology. 2019;10:1900.

84. Huang C, Liu WJ, Xu W, Jin T, Zhao Y, Song J, et al. A bat-derived putative cross-family recombinant coronavirus with a reovirus gene. PLOS Pathogens. 2016 Sep 27;12(9):e1005883-.

85. Xu J, Hu J, Wang J, Han Y, Hu Y, Wen J, et al. Genome organization of the SARS-CoV. Genomics, Proteomics & Bioinformatics. 2003;1(3):226–35.

86. Kim D, Lee JY, Yang JS, Kim JW, Kim VN, Chang H. The architecture of SARS-CoV-2 transcriptome. Cell. 2020;181(4):914-921.e10.

87. Li F. Receptor recognition and cross-species infections of SARS coronavirus. Antiviral Research. 2013;100(1):246–54.

88. Joffrin L, Goodman SM, Wilkinson DA, Ramasindrazana B, Lagadec E, Gomard Y, et al. Bat coronavirus phylogeography in the Western Indian Ocean. Scientific Reports [Internet]. 2020;10(1):6873. Available from: https://doi.org/10.1038/s41598-020-63799-7

89. Razanajatovo NH, Nomenjanahary LA, Wilkinson DA, Razafimanahaka JH, Goodman SM, Jenkins RK, et al. Detection of new genetic variants of Betacoronaviruses in Endemic Frugivorous Bats of Madagascar. Virology Journal [Internet]. 2015;12(1):42. Available from: https://doi.org/10.1186/s12985-015-0271-y

90. Haddad D, John SE, Mohammad A, Hammad MM, Hebbar P, Channanath A, et al. SARS-CoV-2: Possible recombination and emergence of potentially more virulent strains. PLOS ONE [Internet]. 2021 May 25;16(5):e0251368-. Available from: https://doi.org/10.1371/journal.pone.0251368

91. Wang H, Pipes L, Nielsen R. Synonymous mutations and the molecular evolution of SARS-CoV-2 origins. Virus Evolution [Internet]. 2021 Jan 20;7(1). Available from: https://doi.org/10.1093/ve/veaa098

92. Zhou H, Chen X, Hu T, Li J, Song H, Liu Y, et al. A novel bat coronavirus reveals natural insertions at the S1/S2 cleavage site of the spike protein and a possible recombinant origin of HCoV-19. bioRxiv [Internet]. 2020 Jan 1;2020.03.02.974139. Available from: http://biorxiv.org/content/early/2020/03/11/2020.03.02.974139.abstract

93. Li X, Giorgi EE, Marichannegowda MH, Foley B, Xiao C, Kong X-P, et al. Emergence of SARS-CoV-2 through recombination and strong purifying selection. Science Advances [Internet]. 2020 Jul 1;6(27):eabb9153. Available from: http://advances.sciencemag.org/content/6/27/eabb9153.abstract

94. Boni MF, Lemey P, Jiang X, Lam TT-Y, Perry BW, Castoe TA, et al. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. Nature Microbiology [Internet]. 2020;5(11):1408–17. Available from: https://doi.org/10.1038/s41564-020-0771-4

95. Zhu Z, Meng K, Meng G. Genomic recombination events may reveal the evolution of coronavirus and the origin of SARS-CoV-2. Scientific Reports [Internet]. 2020;10(1):21617. Available from: https://doi.org/10.1038/s41598-020-78703-6

96. Graham RL, Baric RS. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. Journal of virology [Internet]. 2009/11/11. 2010 Apr;84(7):3134–46. Available from: https://pubmed.ncbi.nlm.nih.gov/19906932

97. Ge X-Y, Wang N, Zhang W, Hu B, Li B, Zhang Y-Z, et al. Coexistence of multiple coronaviruses in several bat colonies in an abandoned mineshaft. Virologica Sinica [Internet]. 2016/02/18. 2016 Feb;31(1):31–40. Available from: https://pubmed.ncbi.nlm.nih.gov/26920708

98. Brook CE, Ranaivoson HC, Broder CC, Cunningham AA, Héraud J-M, Peel AJ, et al. Disentangling serology to elucidate henipa- and filovirus transmission in Madagascar fruit bats. Journal of Animal Ecology [Internet]. 2019 Jul 1;88(7):1001–16. Available from: https://doi.org/10.1111/1365-2656.12985

99. Huang C, Liu WJ, Xu W, Jin T, Zhao Y, Song J, et al. A bat-derived putative cross-family recombinant coronavirus with a reovirus gene. PLOS Pathogens [Internet]. 2016 Sep 27;12(9):e1005883-. Available from: https://doi.org/10.1371/journal.ppat.1005883

100. Razanajatovo NH, Nomenjanahary LA, Wilkinson DA, Razafimanahaka JH, Goodman SM, Jenkins RK, et al. Detection of new genetic variants of Betacoronaviruses in endemic frugivorous bats of Madagascar. Virology Journal [Internet]. 2015;12(1):42. Available from: https://doi.org/10.1186/s12985-015-0271-y

101. Paskey AC, Ng JHJ, Rice GK, Chia WN, Philipson CW, Foo RJH, et al. Detection of recombinant Rousettus Bat Coronavirus GCCDC1 in Lesser Dawn Bats (Eonycteris spelaea) in Singapore. Viruses [Internet]. 2020 May 14;12(5):539. Available from: https://pubmed.ncbi.nlm.nih.gov/32422932

102. Han Y, Du J, Su H, Zhang J, Zhu G, Zhang S, et al. Identification of diverse bat Alphacoronaviruses and Betacoronaviruses in China provides new insights into the evolution and origin of coronavirus-related diseases. Frontiers in Microbiology [Internet]. 2019;10:1900. Available from: https://www.frontiersin.org/article/10.3389/fmicb.2019.01900

103. Yang Y, Yan W, Hall AB, Jiang X. Characterizing transcriptional regulatory sequences in coronaviruses and their role in recombination. Molecular Biology and Evolution [Internet]. 2021 Apr 1;38(4):1241–8. Available from: https://doi.org/10.1093/molbev/msaa281

104. Li X, Cheng Z, Wang F, Chang J, Zhao Q, Zhou H, et al. A negative feedback model to explain regulation of SARS-CoV-2 replication and transcription. Frontiers in Genetics [Internet]. 2021;12:202. Available from: https://www.frontiersin.org/article/10.3389/fgene.2021.641445

105. Pyrc K, Jebbink MF, Berkhout B, van der Hoek L. Genome structure and transcriptional regulation of human coronavirus NL63. Virology Journal [Internet]. 2004;1(1):7. Available from: https://doi.org/10.1186/1743-422X-1-7

106. Woo PCY, Huang Y, Lau SKP, Yuen K-Y. Coronavirus genomics and bioinformatics analysis. Viruses [Internet]. 2010/08/24. 2010 Aug;2(8):1804–20. Available from: https://pubmed.ncbi.nlm.nih.gov/21994708

107. Mélade J, Wieseke N, Ramasindrazana B, Flores O, Lagadec E, Gomard Y, et al. An eco-epidemiological study of Morbilli-related paramyxovirus infection in Madagascar bats reveals host-switching as the dominant macro-evolutionary mechanism. Scientific Reports [Internet]. 2016;6(1):23752. Available from: https://doi.org/10.1038/srep23752

108. Lebarbenchon C, Ramasindrazana B, Joffrin L, Bos S, Lagadec E, le Minter G, et al. Astroviruses in bats, Madagascar. Emerging microbes & infections [Internet]. 2017 Jun 21;6(6):e58–e58. Available from: https://pubmed.ncbi.nlm.nih.gov/28634357

109. A WD, Julien M, Muriel D, Beza R, Voahangy S, Erwan L, et al. Highly diverse morbillivirus-related paramyxoviruses in wild fauna of the southwestern Indian Ocean Islands: Evidence of exchange between introduced and endemic small mammals. Journal of Virology [Internet]. 2014 Aug 1;88(15):8268–77. Available from: https://doi.org/10.1128/JVI.01211-14

110. Wells HL, Letko M, Lasso G, Ssebide B, Nziza J, Byarugaba DK, et al. The evolutionary history of ACE2 usage within the coronavirus subgenus Sarbecovirus. Virus Evolution [Internet]. 2021 Jan 20;7(1). Available from: https://doi.org/10.1093/ve/veab007

111. Lam TT-Y, Jia N, Zhang Y-W, Shum MH-H, Jiang J-F, Zhu H-C, et al. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. Nature [Internet]. 2020;583(7815):282–5. Available from: https://doi.org/10.1038/s41586-020-2169-0

112. Olival KJ, Cryan PM, Amman BR, Baric RS, Blehert DS, Brook CE, et al. Possibility for reverse zoonotic transmission of SARS-CoV-2 to free-ranging wildlife: A case study of bats. PLOS Pathogens [Internet]. 2020 Sep 3;16(9):e1008758-. Available from: https://doi.org/10.1371/journal.ppat.1008758

113. World Health Organization. Coronavirus in Madagascar . World Health Organization. 2021.

114. Rabalski L, Kosinski M, Mazur-Panasiuk N, Szewczyk B, Bienkowska-Szewczyk K, Kant R, et al. Zoonotic spillover of SARS-CoV-2: mink-adapted virus in humans. bioRxiv [Internet]. 2021 Jan 1;2021.03.05.433713. Available from: http://biorxiv.org/content/early/2021/03/05/2021.03.05.433713.abstract

115. Plowright RK, Peel AJ, Streicker DG, Gilbert AT, McCallum H, Wood J, et al. Transmission or within-host dynamics driving pulses of zoonotic viruses in reservoir–host populations. PLOS Neglected Tropical Diseases [Internet]. 2016 Aug 4;10(8):e0004796-. Available from: https://doi.org/10.1371/journal.pntd.0004796

116. Rocha R, Aziz SA, Brook CE, Carvalho WD, Cooper-Bohannon R, Frick WF, et al. Bat conservation and zoonotic disease risk: a research agenda to prevent misguided persecution in the aftermath of COVID-19. Animal Conservation [Internet]. 2021 Jun 1;24(3):303–7. Available from: https://doi.org/10.1111/acv.12636

117. Kofoky A, Andriafidison D, Ratrimomanarivo F, Razafimanahaka HJ, Rakotondravony D, Racey PA, et al. Habitat use, roost selection and conservation of bats in Tsingy De Bemaraha National Park, Madagascar. Biodiversity and Conservation [Internet]. 2007;16(4):1039–53. Available from: https://doi.org/10.1007/s10531-006-9059-0

118. B Jenkins RK, Racey PA. Bats as bushmeat in Madagascar [Internet]. Available from: http://www.mwc-info.net/en/services/journal.htm

**Figure Legends**

**Figure 1.** **(A)** Map of sampling sites for *P. rufus*, *E. dupreanum*, and *R. madagascariensis* in the Districts of Moramanga and Manjakandriana, Madagascar. Pie charts correspond to coronavirus prevalence in juveniles vs. adults across all three species: 3/15 (20%) vs. 1/29 (3.45%) for *P. rufus,*  5/13 (38.46%) vs. 11/132 (8.33%) for *E. dupreanum*, and 0/13 (0%) vs. 8/83 (9.64%) for *R. madagascaiensis*. Pie circle size corresponds to sample size on a log-10 scale. **(B)** Seasonal variation in adult vs. juvenile CoV prevalence by species, from sites depicted in (A).

**Figure 2.** Genome structure of novel *Nobecoviruses* derived from *P. rufus* and *R. madagascariensis* fruit bats. TRS locations are highlighted by red arrows, and genes are distinguished by color, with orange corresponding to ORF1a and ORF1b, and various shades of blue to structural proteins S, E, M, and N. Accessory genes NS3, NS7a, NS7b, and NS7c (*R. madagascariensis* genomes only) are depicted in powder blue.

**Figure 3. (A)** Maximum Likelihood phylogeny of full genome *Betacoronavirus* sequences, (RAxML-NG, GTR+I+G4) and **(B)** RdRp phylogeny of a 259bp fragment of *Betacoronavirus* Orf1b (RAxML-NG, TVM+I+G4). Bootstrap support values computed using Felsenstein’s method (80) are visualized on tree branches. In both (A) and (B), novel Madagascar sequences are highlighted in yellow, and tip points are colored by *Betacoronavirus* subgenus, corresponding to the legend. Tip shape indicates whether the virus is derived from a bat (triangle) or non-bat (circle) host. Both trees are rooted in turkey *Gammacoronavirus,* accession number NC\_010800.

**Figure 4.** Maximum Likelihood amino acid phylogenies corresponding to translated sequences of the **(A)** spike, **(B)** envelope, **(C)** matrix, and **(D)** nucleocapsid *Betacoronavirus* proteins. All phylogenies were computed in RAxML-NG, using respective amino acid substitution models (A) WAG+I+G4+F, (B) LG+G4, (C) LG+I+G4, and (D) LG+I+G4+F. Bootstrap support values computed using Felsenstein’s method (80) are visualized on tree branches. In (A-D) novel Madagascar sequences are highlighted in yellow, and tip points are colored by *Betacoronavirus* subgenus, corresponding to the legend. Tip shape indicates whether the virus is derived from a bat (triangle) or non-bat (circle) host. Both trees are rooted in turkey *Gammacoronavirus,* accession number NC\_010800.

**Figure 5. (A)** Similarity and **(B)** Bootscan plots computed in SimPlot (v.3.5.1), using a query sequence of *P. rufus* (top panels) and *R. madagascariensis* (bottom panels)*-*derived *Nobecovirus* sequences, with reference sequences corresponding to HKU9 (EF065514-EF065516, HM211098-HM211100, MG693170, NC\_009021, MG762674) and *Eidolon helvum* Africa-derived (MG693169, MG693171-MG693172, NC\_048212 ) *Nobecovirus* lineages. Line color indicates similary (A) and bootscan grouping (B) of the query sequence with the corresponding *Nobecovirus* genotype, along disparate regions of the CoV genome, as indicated by the colored bar at the bottom of each plot. All analyses were carried out using a window size of 200bp and a step size of 20bp.

**Table 1. Putative Transcription Regulatory Sequences in novel *Nobecoviruses* from Madagascar fruit bats.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Coronavirus** | **ORF** | **TRS location (nt)** | **Leader TRS (nt)** | **TRS region** | **Intergenic TRS** | **Distance from TRS to AUG (nt)** |
| P\_rufus\_AMB140 | ORF1ab | 65-70 | UGAA | ACGAAC | UUAAU | 21 |
|  | S | 20667-20672 | GUGA | ACGAAC | UUGUC | 69 |
|  | NS3 | 24593-24598 | AAAG | ACGAGC | UUA**AUG** | 3 |
|  | E | 25240-25245 | UUUA | ACGAAC | GUC**AUG** | 3 |
|  | M | 25451-25456 | UUGA | ACGAAC | AACAA | 14 |
|  | N | 26170-26175 | UUGA | ACGAAA | UUAAA | 6 |
|  | NS7a | 27651-27656 | AAGC | AGCAAC | AGAAA | 70 |
|  | NS7b | 28441-28446 | UUGA | ACGAAG | **AUG** | 0 |
| R\_madagascariensis\_MIZ178 | ORF1ab | 55-60 | GUUG | AAGAAC | UUUAA | 7 |
|  | S | 20790-20795 | UUGA | ACGAAC | UAAAA | 13 |
|  | NS3 | 24575-24580 | UUGA | ACGAAC | UUGUU | 20 |
|  | E | 25293-25298 | GUAA | ACGAAC | UGUAU**AUG** | 5 |
|  | M | 25508-25513 | GAUG | UCGAAC | UAUA**AUG** | 4 |
|  | N | 26255-26260 | UUGA | ACGAAC | AACAA | 17 |
|  | NS7a | 27674-27679 | UUGA | ACGAAC | CAAUU**AUG** | 5 |
|  | NS7b | 28046-28051 | UUGA | ACGAAC | **AUG** | 0 |
|  | NS7c | 28256-28261 | UUUU | ATCAAC | CCGGG | 27 |
| R\_madagascariensis\_MIZ240 | ORF1ab | 119-124 | UUGA | ACGAAC | CU**AUG** | 2 |
|  | S | 20854-20859 | UUGA | ACGAAC | UAAAA | 13 |
|  | NS3 | 24666-24671 | UUGA | ACGAAC | UUGUU | 20 |
|  | E | 25384-25389 | GUAA | ACGAAC | UGUAUA | 6 |
|  | M | 25599-25604 | GAUG | UCGAAC | UAUAA | 8 |
|  | N | 26346-26351 | UUGA | ACGAAC | AACAA | 154 |
|  | NS7a | 27771-27776 | UUGA | ACGAAC | CAAUU**AUG** | 5 |
|  | NS7b | 28216-28221 | UUGA | ACGAAC | **AUG** | 0 |
|  | NS7c | 28353-28358 | GUCG | AGAAAG | AGACC | 59 |

**Table S1. BLAST identity of novel Madagascar *Nobecoviruses* to publicly available reference genomes**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Virus** | **Nucleotide or amino acid** | **β-HKU9**  **(MG762674)** | **β--GCCDC1**  **(NC\_030886)** | **β-BtRt-BetaCoV/GX2018**  **(MK211379)** | **β-Eidolon\_helvum/Cameroon/2013**  **(NC\_048212)** | **Top BLAST hit** | **Accession number of top BLAST hit** |
| P\_rufus\_AMB130 | Genome\* | 73.15\* | 72.87\* | 73.28\* | 73.54\* | Bat coronavirus HKU9, complete genome | EF065514 |
|  | ORF1a | 47.67 | 53.7 | 53.77 | 48.25 | ORF1a [Bat coronavirus] | AWV67046 |
|  | ORF1b | 70.08 | 75.2 | 74.68 | 76.10 | ORF1ab polyprotein [Bat coronavirus] | YP\_009824989 |
|  | S | 45.59 | 47.3 | 46.7 | 46.47 | spike glycoprotein [Bat coronavirus] | QEH60463 |
|  | NS3 | 40.38 | 41.18 | 4.6† | 41.67 | hypothetical protein [Bat coronavirus HKU9] | ADM33567 |
|  | E | 50.68 | 47.95 | 50.63 | 46.58 | envelope protein [Eidolon bat coronavirus/Kenya/KY24/2006 | ADX59468 |
|  | M | 58.56 | 59.73 | 60.55 | 62.44 | Membrane glycoprotein [Bat coronavirus HKU9] | ABN10930 |
|  | N | 50 | 50.61 | 50.84 | 48.27 | N protein [Rousettus bat coronavirus HKU9] | AVP25400 |
|  | NS7a† | 3.3† | 2.1† | 1.9† | 2.2† | CMRF35-like molecule 1 viral protein [Murine norovirus GV/CR10/2005/USA]† | 5OR7† |
|  | NS7b | 40.28 | 0.1† | 63.64 | 40 | NS7b [Rousettus bat coronavirus HKU9] | AVP25412 |
| R\_madagascariensis\_MIZ178 | Genome\* | 77\* | 75.08\* | 75.88\* | 95.15\* | Bat coronavirus isolate CMR900 | MG693169 |
|  | ORF1a | 60.59 | 59.36 | 59.84 | 96.03 | ORF1a [Bat coronavirus] | AWV67062 |
|  | ORF1b | 82.02 | 81.09 | 82.54 | 99.02 | ORF1ab polyprotein [Bat coronavirus] | YP\_009824989 |
|  | S | 49.84 | 50.86 | 51.23 | 66.61 | Spike protein [Bat coronavirus] | AWV67064 |
|  | NS3 | 46.93 | 44.44 | 39.02 | 89.32 | ORF3 protein [Eidolon bat coronavirus Kenya/KY24/2006] | ADX59467 |
|  | E | 58.11 | 55.41 | 64.86 | 92 | envelope protein [Eidolon bat coronavirus/Kenya/KY24/2006] | ADX59468 |
|  | M | 70.70 | 70.75 | 71.09 | 91.98 | membrane protein [Bat coronavirus] | YP\_009824992 |
|  | N | 61.19 | 59.25 | 63.01 | 86.6 | capsid [Bat coronavirus] | AWV67051 |
|  | NS7a | 55.65 | 25.89 | 1.5† | 8.7† | hypothetical protein ORFx [Bat coronavirus] | AWV67068 |
|  | NS7b | 0.1† | 1.3† | 46.67 | 0.8† | hypothetical protein [Bat coronavirus HKU9] | ADM33571 |
|  | NS7c | None | 28.57 | 0† | 45.89 | hypothetical protein ORFy [Eidolon bat coronavirus/Kenya/KY24/2006] | ADX59472 |
| R\_madagascariensis\_MIZ240 | Genome\* | 77.06 | 75.25 | 75.99 | 95.31 | Bat coronavirus isolate CMR891-892 | MG693171 |
|  | ORF1a | 60.54 | 59.41 | 59.77 | 96.15 | ORF1a [Bat coronavirus] | AWV67062 |
|  | ORF1b | 81.91 | 81.12 | 82.28 | 99.13 | ORF1ab polyprotein [Bat coronavirus] | YP\_009824989 |
|  | S | 51.05 | 50.6 | 51.37 | 66.61 | Spike protein [Bat coronavirus] | AWV67064 |
|  | NS3 | 45.81 | 44.44 | 4.7† | 89.74 | ORF3 [Bat coronavirus] | AWV67065 |
|  | E | 58.11 | 55.41 | 64.86 | 92 | envelope protein [Eidolon bat coronavirus/Kenya/KY24/2006] | ADX59468 |
|  | M | 71.43 | 71.23 | 71.29 | 91.98 | membrane protein [Bat coronavirus] | YP\_009824992 |
|  | N | 61.62 | 61.73 | 66.99 | 87.23 | nucleocapsid protein [Eidolon bat coronavirus/Kenya/KY24/2006] | ADX59470 |
|  | NS7a | 53.28 | 2.4† | 0.4† | 1.5† | hypothetical protein ORFx [Bat coronavirus] | AWV67068 |
|  | NS7b | 0† | 2† | 0.5† | 2† | hypothetical protein [Bat coronavirus HKU9] | ADM33571 |
|  | NS7c | None | 28.57 | 0† | 45.89 | hypothetical protein ORFy [Eidolon bat coronavirus/Kenya/KY24/2006] | ADX59472 |