**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.* These novel *Nobecoviruses* 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 both viruses*.* Given the frequency with which coronaviruses, including Nobecoviruses, are known to recombine, these findings emphasize the need for more extensive coronavirus surveillance among wild bats in Africa to document the availability of viral sequences capable of infecting human 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. As bats are consumed widely across the island for subsistence, understanding the landscape of potentially zoonotic coronavirus circulation will be essential to mitigating 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 instead of bats (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), partly 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). Sarbecoviruses which cluster phylogenetically adjacent to ACE2-using lineages have been recently described in Kenyan *Rhinolophid* bats (36,37), highlighting the need for more intensive coronavirus surveillance in Africa. 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. Indeed, recombination has been implicated in many cross-species coronavirus emergence events (including zoonoses) (26,38–41), emphasizing the importance of widespread surveillance in characterizing the landscape of future zoonotic risks. Several factors, which have been reviewed at length elsewhere (31,42,43), contribute to the CoV affinity for recombination, including a large genome size supported by a unique proofreading mechanism in the CoV 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 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 across the island (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* (29,55). Previous studies have demonstrated that this latter *Betacoronavirus* RNA clusters with subgenus *Nobecovirus* (29,55); *Nobecoviruses* have been previously described in Pterodidae fruit bats across Asia and in both East (Kenya) and West (Cameroon) 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), recombination is a distinct possibility in the Madagascar CoV system. Though no *Rhinolophus* spp. bats, the typical host for ACE2-using *Sarbecoviruses*, inhabit Madagascar, the island is home to several species of bat in family Hipposideridae, which host the closely-related 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 tourist visitation sites (63–66). In addition to natural CoV diversity circulating in Malagasy bats, the Embecoviruses, HCoV-OC43 and HCoV-HKU1, and, more recently, the Sarbecovirus, SARS-CoV-2, are known to circulate widely among human hosts in Madagascar (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, assessing their past and future capacity for recombination and relatedness to previously described Nobecoviruses from Asia and other parts of Africa.

**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 (302), throat (143), and urine (196) swabs 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 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 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 of “bat” compiled from all publicly full-length bat genomes in GenBank. 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 blast against a reference database constructed from all full-length reference sequences for *Alpha-* and *Betacoronavirus* available in GenBank. 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 GenBank (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 GenBank (74). We queried identity to reference sequences for four previously described *Nobecovirus* strains (accession numbers: xx (HKU9), xx (GCCDC1), xx (GX2018), and xx (*Eidolon helvum* sequences), as well as to the top BLAST hit overall. In one instance 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 GenBank 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 GenBank records, corresponding to all available full genome sequences with bat hosts under GenBank 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\_010800.1). 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 GenBank. Finally, 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 Gammacoronvirus 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 derived from annotation tracks reported in GenBank or manual annotation 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 trees were aligned in MAFFT v7 (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. Sequences were first aligned in MAFFT v7 (76,77) using default parameter values with full genome sequences corresponding to two disparate *Nobecovirus* genotypes, the HKU9 (EF065514-EF065516, HM211098-HM211100, MG693170, NC\_009021, MG762674) and the *Eidolon helvum* genotypes (MG693169, MG693171-MG693172, NC\_048212). *Nobecovirus* sequences corresponding to the GCCDC1 (21,60) and GX2018/BatCoV92 (58,83) genotypes were left out of recombination analyses due to the presence of inserted genes and/or genetic material upstream from N in the corresponding genomes, which interfered with the alignment.

After alignment, genomes were analyzed for recombination in the program SimPlot (version 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* clades as references, and the corresponding Madagascar sequence as the alternative. Bootscan analyses were also 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*, one from *E. dupreanum* were submitted to GenBank 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%) *P. rufus* samples*,* 16/145 (11.03%) *E. dupreanum* samples, and 8/96 (8.33%) *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 (21,60).

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 (83–85). 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 (**Table 1**), 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, highlighting the dynamic nature of the CoV genome in this region (the site of previously described orthoreovirus recombination in genotype GCCDC1 (84)). 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 (86)), which could be indicative of variation in gene expression across individual bats and/or species.

All Madagascar *Nobecovirus* sequences appeared highly divergent, with BLAST analysis of the full genome *P. rufus Nobecovirus* demonstrating the highest identity at only XX% to previously described *Nobecoviruses* in the HKU9 lineage circulating in Asia and *R. madagascariensis* genomes demonstrating the highest identity at only XX% to *Eidolon helvum Nobecovirus* lineages circulating in Africa (**Table S1**). 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 (XX% identity for *P. rufus* to HKU9and XX% 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, N, and NS7 regions of their genomes, showing only XX-XX% identity to the closest hits across these regions.

*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 (57) (**Figure 3A**). We observed distinct groupings of five main *Nobecovirus* lineages: (a) the largely Asian-derived HKU9 sequences, (b) the African *E. helvum* sequences (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* genomes 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 (55). Characterization of the full length genome of this virus will be needed to clarify whether it represents a genetic variant of or 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 *Betecoronavirus* 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 the E gene phylogeny grouped the *P. rufus* sequence adjacent to the one Cameroonian-derived 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 S1 subunit (first half) of the spike protein, which includes the receptor binding domain that mediates viral entry into host cells (87). 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**

* 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 Africa(73). 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 response(83). 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.

Contribution to the Field Statement:

**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:**

**Data Availability Statement:**

**References**

Figure Legends

**Fig 1**: Map of sampling sites for P. rufus, E. dupreanum, and R. madagascariensis. Circles are in log scale and sorted by CoV negative or positive and adults or juvenile, CoV prevalence in P. rufus, E. dupreanum, and R. madagascariensis over time

**Fig 2:** Genome structure of isolated full genomes, TRS table in word format

**Fig 3:** Full genome+RdRp phylogeny

**Fig 4:** Simplot+bootscan to look for recombination

**Supplementary figs:** BLAST table, phylogenies of N, S, M, E

**Table 1:** TRS locations

**Table 2:** BLAST results

**References**

1. Banerjee, A., Kulcsar, K., Misra, V., Frieman, M. & Mossman, K. Bats and Coronaviruses. *Viruses* **11**, 41 (2019).

2. Hu, B., Ge, X., Wang, L.-F. & Shi, Z. Bat origin of human coronaviruses. *Virology journal* **12**, 221 (2015).

3. Wu, F. *et al.* A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).

4. Ravelomanantsoa, N. A. F. *et al.* The zoonotic potential of bat-borne coronaviruses. *Emerging Topics in Life Sciences* **4**, (2020).

5. Drexler, J. F. *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* **84**, 11336–11349 (2010).

6. Hu, 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* **13**, e1006698–e1006698 (2017).

7. Anthony, S. J. *et al.* Further Evidence for Bats as the Evolutionary Source of Middle East Respiratory Syndrome Coronavirus. *mBio* **8**, e00373-17 (2017).

8. Woo, P. C., Lau, S. K., Li, K. S., Tsang, A. K. & 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* **1**, e35–e35 (2012).

9. Corman, V. M. *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* **88**, 11297–11303 (2014).

10. Razanajatovo, N. H. *et al.* Detection of new genetic variants of Betacoronaviruses in Endemic Frugivorous Bats of Madagascar. *Virology Journal* **12**, 42 (2015).

11. P, L. S. K. *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* **84**, 11385–11394 (2010).

12. Frutos, R., Serra-Cobo, J., Pinault, L., Lopez Roig, M. & Devaux, C. A. Emergence of Bat-Related Betacoronaviruses: Hazard and Risks. *Frontiers in Microbiology* **12**, 437 (2021).

13. Chen, S.-C., Olsthoorn, R. C. L. & 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* **7**, (2021).

14. Zhou, Z., Qiu, Y. & Ge, X. The taxonomy, host range and pathogenicity of coronaviruses and other viruses in the Nidovirales order. *Animal Diseases* **1**, 5 (2021).

15. Forni, D., Cagliani, R. & Sironi, M. Recombination and Positive Selection Differentially Shaped the Diversity of Betacoronavirus Subgenera. *Viruses* **12**, 1313 (2020).

16. Llanes, A. *et al.* Betacoronavirus Genomes: How Genomic Information has been Used to Deal with Past Outbreaks and the COVID-19 Pandemic. *International journal of molecular sciences* **21**, 4546 (2020).

17. Li, W. *et al.* Bats Are Natural Reservoirs of SARS-Like Coronaviruses. *Science* **310**, 676 (2005).

18. Lam, T. T.-Y. *et al.* Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature* **583**, 282–285 (2020).

19. Hul, V. *et al.* A novel SARS-CoV-2 related coronavirus in bats from Cambodia. *bioRxiv* 2021.01.26.428212 (2021) doi:10.1101/2021.01.26.428212.

20. Paskey, A. C. *et al.* Detection of Recombinant Rousettus Bat Coronavirus GCCDC1 in Lesser Dawn Bats (Eonycteris spelaea) in Singapore. *Viruses* **12**, 539 (2020).

21. Valitutto, M. T. *et al.* Detection of novel coronaviruses in bats in Myanmar. *PLOS ONE* **15**, e0230802- (2020).

22. Lau, S. K. P. *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* **84**, 2808–2819 (2010).

23. Latinne, A. *et al.* Origin and cross-species transmission of bat coronaviruses in China. *Nature Communications* **11**, 4235 (2020).

24. Wacharapluesadee, S. *et al.* Diversity of coronavirus in bats from Eastern Thailand. *Virology Journal* **12**, 57 (2015).

25. Ying, T. *et al.* Surveillance of Bat Coronaviruses in Kenya Identifies Relatives of Human Coronaviruses NL63 and 229E and Their Recombination History. *Journal of Virology* **91**, e01953-16 (2021).

26. Montecino-Latorre, D. *et al.* Reproduction of East-African bats may guide risk mitigation for coronavirus spillover. *One Health Outlook* **2**, 2 (2020).

27. Tong, S. *et al.* Detection of novel SARS-like and other coronaviruses in bats from Kenya. *Emerging infectious diseases* **15**, 482–485 (2009).

28. Joffrin, L. *et al.* Bat coronavirus phylogeography in the Western Indian Ocean. *Scientific Reports* **10**, 6873 (2020).

29. Razanajatovo, N. H. *et al.* Detection of new genetic variants of Betacoronaviruses in Endemic Frugivorous Bats of Madagascar. *Virology Journal* **12**, 42 (2015).

30. Anthony, S. J. *et al.* Coronaviruses in bats from Mexico. *The Journal of general virology* **94**, 1028–1038 (2013).

31. Frutos, R., Serra-Cobo, J., Pinault, L., Lopez Roig, M. & Devaux, C. A. Emergence of Bat-Related Betacoronaviruses: Hazard and Risks. *Frontiers in Microbiology* **12**, 437 (2021).

32. Markotter, W., Coertse, J., de Vries, L., Geldenhuys, M. & Mortlock, M. Bat-borne viruses in Africa: a critical review. *Journal of zoology (London, England : 1987)* 10.1111/jzo.12769 (2020) doi:10.1111/jzo.12769.

33. Motayo, B. O., Oluwasemowo, O. O. & Akinduti, P. A. Evolutionary Dynamics And Geographic Dispersal Of Beta Coronaviruses In African Bats. *bioRxiv* 2020.05.14.056085 (2020) doi:10.1101/2020.05.14.056085.

34. Plowright, R. K., Becker, D. J., McCallum, H. & Manlove, K. R. Sampling to elucidate the dynamics of infections in reservoir hosts. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **374**, 20180336 (2019).

35. Becker, D. J., Crowley, D. E., Washburne, A. D. & Plowright, R. K. Temporal and spatial limitations in global surveillance for bat filoviruses and henipaviruses. *Biology Letters* **15**, 20190423 (2019).

36. Washburne, A. D. *et al.* Taxonomic patterns in the zoonotic potential of mammalian viruses. *PeerJ* **6**, e5979–e5979 (2018).

37. Plowright, R. K. *et al.* Transmission or Within-Host Dynamics Driving Pulses of Zoonotic Viruses in Reservoir–Host Populations. *PLOS Neglected Tropical Diseases* **10**, e0004796- (2016).

38. Banerjee, A. *et al.* Novel Insights Into Immune Systems of Bats. *Frontiers in Immunology* **11**, 26 (2020).

39. Rocha, R. *et al.* Bat conservation and zoonotic disease risk: a research agenda to prevent misguided persecution in the aftermath of COVID-19. *Animal Conservation* **24**, 303–307 (2021).

40. B Jenkins, R. K. & Racey, P. A. *Bats as bushmeat in Madagascar*. http://www.mwc-info.net/en/services/journal.htm.

41. Becker, D. J. *et al.* Optimizing predictive models to prioritize viral discovery in zoonotic reservoirs. *bioRxiv* 2020.05.22.111344 (2021) doi:10.1101/2020.05.22.111344.

42. Haddad, D. *et al.* SARS-CoV-2: Possible recombination and emergence of potentially more virulent strains. *PLOS ONE* **16**, e0251368- (2021).

43. Olival, K. J. *et al.* Possibility for reverse zoonotic transmission of SARS-CoV-2 to free-ranging wildlife: A case study of bats. *PLOS Pathogens* **16**, e1008758- (2020).

44. Kumakamba, C. *et al.* Coronavirus surveillance in Congo basin wildlife detects RNA of multiple species circulating in bats and rodents. *bioRxiv* 2020.07.20.211664 (2020) doi:10.1101/2020.07.20.211664.

45. Ar Gouilh, M. *et al.* SARS-CoV related Betacoronavirus and diverse Alphacoronavirus members found in western old-world. *Virology* **517**, 88–97 (2018).

46. Brook, C. E. *et al.* Disentangling serology to elucidate henipa- and filovirus transmission in Madagascar fruit bats. *Journal of Animal Ecology* **88**, 1001–1016 (2019).

47. Kofoky, A. *et al.* Habitat Use, Roost Selection and Conservation of Bats in Tsingy De Bemaraha National Park, Madagascar. *Biodiversity and Conservation* **16**, 1039–1053 (2007).

48. Rocha, R. *et al.* Human-Bat Interactions in Rural Southwestern Madagascar through a Biocultural Lens. *Journal of Ethnobiology* **41**, 53–69 (2021).

49. Olival, K. J. *et al.* Host and viral traits predict zoonotic spillover from mammals. *Nature* **546**, 646–650 (2017).

50. Ge, X.-Y. *et al.* Coexistence of multiple coronaviruses in several bat colonies in an abandoned mineshaft. *Virologica Sinica* **31**, 31–40 (2016).

51. Wang, H., Pipes, L. & Nielsen, R. Synonymous mutations and the molecular evolution of SARS-CoV-2 origins. *Virus Evolution* **7**, (2021).

52. Zhou, H. *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* 2020.03.02.974139 (2020) doi:10.1101/2020.03.02.974139.

53. Li, X. *et al.* Emergence of SARS-CoV-2 through recombination and strong purifying selection. *Science Advances* **6**, eabb9153 (2020).

54. Boni, M. F. *et al.* Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nature Microbiology* **5**, 1408–1417 (2020).

55. Graham, R. L. & Baric, R. S. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. *Journal of virology* **84**, 3134–3146 (2010).

56. Wells, H. L. *et al.* The evolutionary history of ACE2 usage within the coronavirus subgenus Sarbecovirus. *Virus Evolution* **7**, (2021).