Detection and characterization of whole genome sequence of novel astrovirus in an endemic Malagasy fruit bat: *Rousettus madagascariensis*

**INTRODUCTION**

Bats (order: Chiroptera) are an extremely diverse mammalian order that in recent years have been posited as reservoirs of many of the world’s most virulent emerging viruses 1–3. This disproportioned representation is thought to be due to a number of reasons, including immune system modifications due to the evolution of flight, long lifespans, and dense roosting aggregations that facilitate intra- and inter- species transmission 1,4,5. Despite their importance as reservoirs, thorough surveillance of the viruses hosted by bats has been restricted to few species and locations.

The island of Madagascar provides an important case study for viral surveillance to test the presence and evolutionary history of bat viruses and investigate their zoonotic potential. Madagascar hosts 46 bat species with nearly 80% endemism, the result of a unique biogeographical history6,7. This isolation has also led to the coevolution of extraordinarily diverse viruses within Malagasy bats8–10. Moreover, high human-bat contact rates due to hunting, particularly of large fruit bats, across the island increase the potential for zoonotic spillover11. Despite this, viruses within Malagasy bats remain understudied.

A key understudied family of viruses are Astroviruses (AstVs) (family: *Astroviridae*). Astroviruses are non-enveloped, positive-sense, single-stranded RNA viruses. Their genome contains a 5’-untranslated region (UTR), three open reading frames (ORFs)—ORF1a, ORF1b, and ORF2—and a 3’UTR with a poly A tail12. Astroviruses infect a remarkable diversity of hosts, grouping broadly into two genera: *Avastrovirus,* which infect avian hosts, and *Mamastrovirus,* which infect mammalian hosts13. This lack of host specificity demonstrates astroviruses efficiency in cross-species transmission, thought to be facilitated by high viral diversity and frequent recombination events12,14–17. There is also evidence of zoonotic transmission, and human astrovirus infections are thought to be responsible for 2-9% of all acute non-bacterial gastrointestinal infections in children12. Despite their clear importance in public health, we have only recently begun to understand the diversity of this family and its evolutionary history.

Astroviruses have been detected with notable presence and remarkable diversity in both old-world and new world bats (*Yinpterochiroptera* and *Yangochiroptera*, respectively), though these values range widely depending on the study design, species, and location18–22. Differing phylogenetic analyses reveal both strong and weak host- and geographic- clustering, with some bat astroviruses grouping more closely with members of the avian *Avastrovirus* genus. The detection of high astrovirus diversity detected within members of the same population of hosts suggests the simultaneous circulation of multiple strains, perhaps driven by the co-roosting of multiple species in dense aggregations1,4,23. These insights, however, have been largely based on single-gene PCR data from few species of bats, limiting the potential for more robust evolutionary analysis. Very recently, metagenomic next-generation (mNGS) sequencing has resulted in the detection of five full-genome bat astroviruses globally. However, limited analysis has been done with these sequences and representation among bat species and geographic location remains lacking.

This paper presents results on (mNGS) detection and characterization of astroviruses sampled from three species of endemic fruit bats from Madagascar*: Pteropus rufus,* *Rousettus madagascariensis,* and *Eidolon dupreanum*. We present a novel full-length genome astrovirus detected in *Rousettus madagascariensis* and characterize its evolutionary history among the broader clade of *Mamastroviruses.* We also utilize a conserved region of this genome to explore the biogeographical history of bat astroviruses from Madagascar and its surrounding landmasses in the Southwest Indian Ocean Region. Through both aims we highlight both the diversity of astroviruses as well as viruses hosted within Malagasy bats.

**METHODS**

**Sample collection and processing**

Astrovirus infections were identified from a dataset of viruses detected in samples from a long-term longitudinal sampling of fruit bats across Madagascar. Details on bat sampling and sequencing protocol can be found in previous work9,10 and on our Github (https://github.com/brooklabteam/Mada-Bat-AstV); here we will only give a brief overview.

Between 2018 – 2019, monthly bat captures were performed at four species-specific locations: Ambakoana roost (-18.513 S, 48.167 E, *Pteropus rufus)*; Angavobe cave (-18.944 S, 47.949 E, *Eidolon dupreanum)*; Angavokely cave (-18.933 S, 47.758 E, *Eidolon dupreanum)*; Maromizaha cave (-18.9623 S, 48.4525 E, *Rousettus madagascariensis).* Bats were identified to species, sex, and age (adult vs juvenile), and throat, fecal, and urine samples were taken.

In total, RNA from 285 fecal, 143 throat, and 196 urine swab samples was extracted in the Virology Unit at the Institut Pasteur de Madagascar using the Zymo Quick DNA/RNA Microprep Plus Kit (Zymo Research, Irvine, CA). Extractions were stored at -80 degrees, then transported on dry ice to the Chan Zuckerburg Biohub (CZID) (San Francisco, CA, USA) for library preparation and metagenomic Next Generation Sequencing (mNGS).

Aliquots of each sample were arrayed into a 384 well plate for input into mNGS library prep. Samples were evaporated using a GeneVac EV-2 (SP Industries, Warminster, PA, USA) to enable miniaturized mGNS library preparation with the NEBNext Ultra II RNA Library Prep Kit (New England Biolands, Beverly, MA, USA). Library preparation was performed per the manufacturer’s instructions, with the following modifications: 25 pg of External RNA Controls Consortium Spike-in mix (ERCCS, Thermo-Fisher) was added to each sample prior to RNA fragmentation; the input RNA mixture was fragmented for 8 min at 94 degrees C prior to reverse transcription; and a total of 14 cycles of PCR with dual-indexed TruSeq adaptors was applied to amplify the resulting individual libraries. Samples were assessed for quality and quantity, then submitted to the Illumina NovaSeq (Illumina, San Diego, CA, USA) for paired-end sequencing (2 x 146 bp). The pipeline used to separate the sequencing output of the individual libraries into FASTQ files of 146bp paired-end reads is available on Github at https://github.com/czbiohub/utilities.

Raw reads from Illumina were host-filtered, quality-filtered, and assembled on the CZID bioinformatics platform (v3.10 NR/NT 2019-12-01)24, using a host background model of “bat” compiled from all publicly available full-length bat genomes in GenBank at the time of sequencing. Samples were marked positive for astrovirus infection if at least two contigs with an average read depth >2 reads/nucleotide were assembled that showed significant nucleotide or protein BLAST alignment(s) (alignment length >100nt/aa and E-value < 0.00001 for nucleotide BLAST/bit score >100 for protein BLAST) to astroviruses present in NCBI NR/NT database (v12-01-2019). Additionally, all non-host contigs assembled in CZID were manually BLASTed against all full-length and protein reference sequences for astroviruses available in NCBI Virus.

**Genome Annotation and Similarity Analysis**

To annotate coding sequences, we downloaded all available bat astrovirus full genomes published on NCBI Virus. We aligned these annotated genomes to our novel genome using MAFFT25 (v7.450) with default parameters in Geneious Prime (v08-18-2022). We then annotated open reading frames and genes in our novel sequence by identifying stop and start codons in regions adjacent to those identified in the homologs.

To investigate similarity to other published sequences, we performed blastn (nucleotide-nucleotide) and blastx (translated nucleotide-protein) searches within the NCBI database and downloaded the hit tables (see supplement). Additionally, we created animo acid and nucleotide similarity plots using the program pySimplot with input alignmented generated using MAFFT25 (v7.450) with default parameters in Geneious Prime (v08-18-2022).

**Phylogenetic Analysis Overview**

To perform phylogenetic analysis, we combined Astrovirus sequences identified in this study with those publicly available on NCBI. We carried out three major phylogenetic analyses, building (a) a full-genome *Mamastrovirus* maximum (ML) likelihood phylogeny, (b) a time-resolved Bayesian phylogeny corresponding to a selection of full genome *Mamastrovirus* sequences available on NCBI Virus, and (c) a *Mamastrovirus* ML phylogeny corresponding to a conserved 410 bp fragment of the RNA-dependent RNA polymerase (RdRp) gene encapsulated in the AstV ORF1b with a focus in the South West Indian Ocean region. Detailed methods for the construction of each phylogeny are available on GitHub (see supplementary information).

**Sequence Compilation**

Our full genome ML phylogeny consisted of one novel full length *Mamastrovirus* sequence, 41 unique NCBI *Mamastrovirus* sequences, and one full length *Avastrovirus* sequence as an outgroup, for a total of 43 sequences. We compiled the NCBI sequences through three queries: 1) all complete RefSeq Genomes under Virus: *Mamastrovirus* (taxid:249588) and Virus: *unclassified Mamastrovirus* (taxid:526119) greater than 6,000 bp (36), 2) *Mamastrovirus* nucleotide genomes under Virus: *Astroviridae* (taxid:39733) and Virus: *unclassified Astroviridae* (taxid:352926) with Host: *Chiroptera (bats)* (taxid:9397) over 6,000 bp (2), and 3) manual searching of *Mamastrovirus* nucleotide genomes >6,000 bp identified in the literature (3).

Our Bayesian timetree consisted of the same set of full length *Mamastrovirus* sequences, removing the *Avastrovirus* outgroup, for a total of 42 sequences.

Our *Mamastrovirus* RdRp ML phylogeny consisted of an overlapping subset of a 410 bp fragment in the center of the RdRp region of our novel full length *Mamastrovirus* sequence, 122 unique NCBI *Mamastrovirus* sequences, and one Avastrovirus RdRp fragment as an outgroup, for a total of 124 sequences. NCBI sequences were restricted to those from bat hosts sampled in the Southwest Indian Ocean (SWIO) region. They were compiled through one query: Virus: Astroviridae (taxid:39733), Host: Chiroptera (taxid:9397), and Geographic Region: Madagascar (64), Mozambique (31), and Reunion (27). Sequences were confirmed to be RdRp fragments via alignment and relevant metadata was confirmed though source literature.

**Alignment and Substitution Model**

Following dataset compilation for each phylogenetic analysis, sequences were aligned via the MAFFT25 (v7.450) algorithm in Geneious Prime (v 2022-08-18) using default parameters. Alignments were visually examined and trimmed to the shortened length sequence in the dataset. We then used Modeltest-NG26 (v0.1.7) to determine the best fit nucleotide substitution for each alignment. All sequences, subsets, and alignments are available in our open-source GitHub repository (see supplementary information).

**Phylogenetic Tree Assembly**

Both the full genome and RdRp ML trees were constructed in RAxML-NG27 (v1.1.0), using the best nucleotide substation model from Modeltest-NG26. Following best practice recommendations in RAxML-NG27, twenty ML inferences were made, followed by bootstrap replicate trees inferred using Felsenstein’s method28. The MRE-based bootstrapping test was performed every 50 replicates, and bootstrapping was terminated when the diagnostic result was below the threshold value. Support values were compiled onto the best-scoring tree.

The Bayesian time tree was built using BEAST229 (v2.6.7), using the best nucleotide substation model from Modeltest-NG26. We used a Bayesian Skyline Coalescent Model with a strict lognormal clock rate with mean 0.001, and population and group size set to 1. Sampling date for each sequence was inferred from NCBI ‘Collection Date’ or though reading source literature; if day was not available the sampling date was set to the 15th of the month listed; if day and month were not available the sampling date was set to July 15th. Markov Chain Monte Carlo (MCMC) chains were run for >700,000,000 iterations and terminated when we identified convergence at ESS values > 200 using TRACER (v1.7), with 10% burn-in. We used TreeAnnotator (v2.6.3) to examine mean posterior densities at each node.

All phylogenies were visualized in RStudio (v2022.07.01), using the package ‘ggtree’30.

**Nucleotide Sequence Accession Numbers**

One annotated full-length genome sequence from *Rousettus madagascariensis* was submitted to NCBI and is available under accession number OQ606244.

**RESULTS**

**Detection**

Astroviruses were detected in all species across all sampled locations. We detected a total of 11 positive astrovirus samples: 4/285 (1.4%) fecal samples, 7/196 urine samples (3.57%), and 0/143 (0%) throat swabs. Although astroviruses are known to transmit fecal-orally, they have been detected in bat urine across multiple studies31. It is possible as well that cross-contamination of fecal/urine samples occurred given the nature of collection.

Detection across all sample types (urine and feces) varied slightly between species, with 1/44 (2.27%) *Pteropus rufus*, 8/145 (5.52%) *Eidolon dupreanum*, and 2/96 (2.08%) *Rousettus madagascariensis* positives detected, but these differences were not significant (χ2=2.104, P=0.3492). Juvenile vs adult prevalence also did not vary significantly: 1/15 (6.66%) vs. 0/29 (0%) for *P. rufus* (χ2=1.941, P=0.1635), 1/13 (7.69%) vs. 7/132 (5.30%) for *E. dupreanum* (χ2=0.122, P=0.7264), and 1/13 (7.69%) vs. 1/83 (1.20%) for *R. madagascariensis* (χ2=2.271, P=0.1318) (Figure 1).

Diagram

Description automatically generated

Figure 1: Map of sampling sites for *P. rufus*, *E. dupreanum*, and *R. madagascariensis* in the districts of Moramanga and Manjakandriana, Madagascar (*P. rufus:* Ambakoana roost; *E. dupreanum*: Angavobe/Angavokely caves; *R. madagascariensis*: Maromizaha cave). Pie charts correspond to astrovirus prevalence in bats across all three species: 1/15 (6.66%) vs. 0/29 (0%) for *P. rufus*, 1/13 (7.69%) vs. 7/132 (5.30%) for *E. dupreanum*, and 1/13 (7.69%) vs. 1/83 (1.20%) for *R. madagascariensis*. Pie circle size corresponds to sample size on a log-10 scale.

**Genome Annotation and Similarity Analysis**

One near-full length Astrovirus contig with high coverage was recovered from a male, juvenile *Rousettus madagascariensis* bat*,* 6,593 bp in length (Figure 2). By aligning this sequence to annotated full genome *Mamastroviruses* from NCBI, we successfully annotated ORF1a, ORF1b, which contains the RdRp region, and ORF2. ORF1a is near-fully represented; a fragment of the 5’ region including the start codon was not captured in sequencing. All other astrovirus detections represented partial fragments and aligned within the ORF2 region. Because this region is not typically targeted in PCR, there were not enough NCBI submissions available for a thorough evolutionary analysis, thus these hits were not analyzed further.

BLAST analysis of the near-full genome against the NCBI database indicated that this novel astrovirus is highly divergent (see supplement for blast tables). blastn resulted in only one hit with a query coverage above 0%, and it was only 5%, with a percent identity of 80.94%. This hit was to a Mamastrovirus RNA-dependent RNA-polymerase (RdRp) from an unknown Chiropteran species sampled in Tanzania. All other top hits were to *Mamastroviruses*, with the vast majority from Porcine hosts. blastx resulted in hits with higher query coverage: the top hit, at 40% query coverage and 43.09% identity was also to a bat astrovirus, in this case the ORF1 region from *Eidolon helvum* sampled in Cameroon. All other hits in the top were to non-chiropteran species including porcine and raccoon dogs.

This divergence is supported by nucleotide and amino acid similarity plots comparing this novel sequence to all full genome bat Astrovirus sequences from NCBI (Figure 2). Using the novel near-full sequence as a query (OQ606244), we see low similarity across the entire genome, with all other bat sequences cluster together at a range of average nucleotide similarity of 28.18% - 40.28% and average amino acid similarity from 23.87% - 44.29% in open-reading frame (ORF) 1a, 35.14% - 55.61% in ORF1b, and 24.48% - 40.73% in ORF2 (Table 1).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| NCBI ID | Host Species | % nucleotide similarity | % ORF1a similarity | % ORF1b similarity | % ORF2 similarity |
| MG693176 | *Eidolon helvum* | 38.84 | 44.29 | 55.61 | 40.73 |
| MN832787 | *Myotis daubentonii* | 39.84 | 23.87 | 51.78 | 24.84 |
| MT734809 | *Myotis daubentonii* | 40.28 | 24.78 | 50.31 | 24.92 |
| MZ218053 | *Myotis daubentonii* | 28.18 | 25.99 | 54.26 | 25.59 |
| MZ218054 | *Myotis yumanensis* | 28.24 | 25.92 | 35.14 | 26.65 |

Table 1: Summary table of similarity plots showing % similarity of each queried astrovirus to novel astrovirus across whole-genome nucleotide sequence and amino acid open-reading frame (ORF) 1a, 1b, and 2.

There are slight differences in the similarity between astroviruses from the queried species. An astrovirus from one of the sampled *Myotis daubentonii* has the highest average nucleotide similarity to the novel astrovirus, while the highest amino acid similarity across all open reading frame is an astrovirus from *Eidolon helvum*, another Pteropodid.

Timeline

Description automatically generated with medium confidence

Figure 2: Genomic structure of Astroviruses. Similarity plots comparing all chiropteran full genome astrovirus sequences to the novel Madagascar bat astrovirus sequence for nucleotide sequences and amino acid sequences. Query sequence is OQ606244, comparison sequences are MG693176 *(Eidolon helvum)*, MZ218054 *(Myotis daubentonii 1),* MZ218053 *(Myotis daubentoniid 2),* MN832787 *(Myotis daubentoniid 3),* MT734809 *(Myotis yumanensis).* Coverage plot showing contig depth for assembled full-genome OQ606244.

**Phylogenetic Analysis**

Diagram, schematic

Description automatically generated

Figure 3: Maximum likelihood phylogeny of full genome *Mamastrovirus* sequences (RAxML-NG, TVM+I+G4). Bootstrap values computed using Felsenstein’s method28 are visualized on tree branches. Tip labels include NCBI taxon ID, strain, host species, location of collection, and year of collection. Tip points are colored by order of mammalian host and shaped by bat or non-bat host with Madagascar sequences labeled with triangles. The novel astrovirus sequence presented in this paper is highlighted in yellow. Tree is rooted by a turkey *Avastrovirus* (NC\_002470) and a divergent Porcine *Mamastrovirus* (NC\_023636). Branch lengths are scaled by nucleotide substitutions per site, corresponding to the scalebar.

***Mamastrovirus* full-genome evolutionary history**

The full-genome maximum likelihood tree resolves two distinct *Mamastrovirus* clades, with a single Porcine astrovirus grouping with the *Avastrovirus* outgroup. No host order demonstrates monophyly, with high support in most cases. Bat astroviruses are a paraphyletic group basal to the clade containing astroviruses from Artiodactylia, Carnivora, and Primates (in this case, humans).Our novel astrovirus groups most closely with a sample from a Cameroonian *Eidolon helvum* and fall out more recently derived than the insectivorous bat astroviruses*.*

Chart

Description automatically generated

Figure 4: Bayesian phylogeny of full genome *Mamastrovirus* sequences generated from >700,000,000 steps under a Bayesian Skyline Coalescent Model (TVM+I+G4). Node color represents mean posterior estimates averaging over all steps with 10% burn-in (see scale bar on left). Tip labels include NCBI taxon ID, strain, host species, sampling location, and year of sample, and are colored by strain, with our novel astrovirus sequence highlighted in yellow.

**Bayesian time-tree**

Diagram, schematic

Description automatically generated

Figure 5: Maximum Likelihood phylogeny of a 410bp fragment of *Mamastrovirus* Orf1b (RdRp) from Southwest Indian Ocean (SWIO) bat hosts (RAxML-NG, TVM+I+G4). Bootstrap values computed using Felsenstein’s method (CITE) and those >30 are visualized on tree branches. Tip labels include NCBI taxon ID, host family, and host species. Tip points are colored by sub-order and shaped by country of origin. The novel astrovirus sequence presented in this paper is highlighted in yellow. Tree is rooted by a Turkey *Avastrovirus* (NCBI taxID: NC\_002470) and a divergent bat *Mamastrovirus* (NCBI taxID: MZ614426). Branch lengths are scaled by nucleotide substitutions per site, as indicated by the scale bar.

**RdRp SWIO phylogeny**

To investigate the biogeographical evolutionary history of bat astroviruses, we built an RdRp tree with all bat *Mamastrovirus* samples from the Southwest Indian Ocean Region. In all, we had representatives from X species across three sampled locations: Madagascar, Mozambique, and Reunion Island.

AstV sequences in the RdRp phylogeny group largely according to host phylogeny over geography of sampling site, a pattern which holds across both order and family across all sampling locations. Almost all Yinpterochiropteran *Mamastrovirus* sequences fall into a monophyletic group, and Yangochiropterans appear in multiple clades basal to this group. Within Yangochiropterans, Molossids and Miniopterids cluster tightly together regardless of location. Molossid samples did not group within the other microbats, and instead fell out in numerous clades basal to all other samples. Paraphyly has been detected in Molossid species32, and thus may promote paraphyly in their viruses through coevolution. This host restriction, however, may be confounded with sampling location, since only Reunion Island Molossids and Madagascar Miniopterids were represented.

There are a few notable exceptions to this pattern. Nycterid (family: Nycteridae) samples fell out in two disparate locations across the phylogeny: one cluster within the Molossids and another as basal to the Yinpterochiropteran clade. Vespertilionids (family: Vespertilionidae), with a sample size of only two, were placed within the Yinpterochiropteran clade, most closely related to Rhinonycterids (family: Rhinocyteridae).

Another is representatives from Vespertilionidae, which group more closely with Yinpterochiroptera than Yangochiroptera.

Within the Yinpteropchiropterans, Rhinocyterid astroviruses mixed with Hipposiderids, both with representatives from Madagascar and Mozambique. Numerous papers have established them as sister families7,32, thus there genetic similarity to each other makes cross-transmission between them more likely.

Pteropodid samples did not group together or fall out in their accepted phylogenetic placement as basal to Yinpterochiropterans7, despite all being sampled from the same endemic Madagascar species, *Rousettus madagascariensis*. All three samples fell out on very different places on the phylogeny—one within Yinpterochiroptera most closely related to a Rhinocyterid sample, one within the Yangochiropteran Miniopterids, and our novel one within the Yangochiropteran Molossids.

weak node support.

**DISCUSSION**

This paper utilizes mNGS sequencing to detect and characterize astroviruses from three endemic species of Malagasy fruit bats, *Pteropus rufus, Rousettus madagascariensis,* and *Eidolon dupreanum.* It characterizes the first near-full genome Astrovirus detected from *Rousettus madagascariensis.*

**Detection**

This work is the first detection of astroviruses from two endemic species of Malagasy fruit bat: *Eidolon dupreanum* and *Pteropus rufus*. Astroviruses have now been detected in 7 species of Malagasy bat.

While detected across all samples, our prevalence was low, in all cases below 8%, however, this level of infection has been described in other bats, particularly Pteropodids33,34. Other studies, however, detect astrovirus infection as high as 51%35, this is appears to be more commonly seen in insectivorous bats.

Seasonality has been detected in some studies, not in others33. Lebarbchenon et al detected astrovirus in 2/41 sampled *Rousettus madagascariensis* (4.8%).

Differences in detection could be due to sampling bias or could reflect variation in AstV infection.

**Genome Annotation and Similarity Analysis**

Our detection of a whole genome allowed for valuable insight into the genetic diversity of this novel astrovirus strain. Similarity analysis of our full genome with all other bat astrovirus full genomes support its strong diversity. This is in line with other studies demonstrating the extraordinary divergence of Madagascar bat viruses, likely from the long isolation the island has experienced9. Blast results were curious: top hit was to bat, but other bat hits were far lower, often below pig hits.

**Full genome phylogeny**

To place our novel astrovirus strain within the evolutionary history of *Mamastrovirus*, we built a full genome maximum likelihood phylogeny using all available bat full genomes and all available reference full genomes from other host species.

Non-monophyly of bats astros

Additionally, bat astroviruses are basal to several human astroviruses. This supports the known propensity for astroviruses to frequently cross species boundaries, as well as supporting a history of zoonotic transfer.

**Bayesian time-tree**

To investigate the timing of the origination of the currently known bat astrovirus strains, we built a Bayesian time tree of the clade containing all known bat astroviruses. This Bayesian tree resolves the same relationships as the maximum likelihood tree, except it changes which clade is basal. The tree estimates a deep evolutionary history for this Mamastrovirus clade, with the MRCA of bat astroviruses diverging ~30,000 years ago, and the MRCA of our novel bat astrovirus diverging from that from *Eidolon helvum* ~15,000 years ago. Considering Madagascar’s isolation from the African continent for ~160 million years and the divergence of the Rousettus genus ~20 million years ago36, this would indicate recent viral genetic exchange between bats from Madagascar and the African continent. Recent publications investigating the divergence times of other viruses from Malagasy bats support this hypothesis9.

**RdRp SWIO phylogeny**

1. **Host restriction vs geography**

Largely, astrovirus samples grouped by phylogenetic relationship of bat family than by sampling location, following the evolutionary separation of Yinpterochiroptera and Yangochiroptera32. This is in contrast to many other astrovirus studies, where host restriction was found to be low or nonexistent8,34,37–40. Islands?

1. The exceptions?

Interestingly, Molossids from Madagascar have been found to have low to no prevalence of astrovirus depending on sampling location8,41. Sampling bias? Cross-spp transmission?

1. Yinp more likely to cross to yang

It appears from this phylogeny that Yinpterochiropteran astroviruses appear more often within Yangochiropteran clades than vice versa. While there are only 3 microbat samples within the megabat cluster, and they all fall out basally, there are 8 megabat samples found within microbat clusters, and they are widely dispersed throughout the phylogeny.

Close phylogenetic relationships between viruses hosted by different species indicates cross-species transmission. In the case of bats, this is promoted by several features of bat behavior and biology, including immune system modifications due to the evolution of flight, long lifespans, and dense roosting aggregations that facilitate intra- and inter- species transmission 1,4,5. Co-roosting! 42

However, the presence of more Yinpterochiropteran virus samples nested within Yangochiropteran virus samples appears to suggest that the likelihood of transmission between the two groups may not be equal.

**Conclusion**

1. **Astros being discovered all over the place, and are being shown to be even more divergent**
2. **Sampling is critically important, particularly given their zoonotic potential**
3. **This dataset, longitudinal study, is a great model for viral surveillance**

Additionally, because new Astrovirus strains are being discovered more rapidly than ever before, we recommend renaming all strains to reference their host organisms.

Much more extensive sampling to increase the number of whole genome bat astroviruses is clearly necessary to resolve relationships.

Because these species and locations are part of a long-term longitudinal study, there is strong future potential to further develop patterns of AstV variation in seasonality, locations, and hosts. However, our consistent detection of astroviruses across bat species and geographic locations lends support to bats as likely reservoirs of AstV’s.

DATA AVAILABLITY STATEMENT

The sequences presented in the study are deposited in NCBI, accession number: XXXX. Detailed methods are available at www.github.com/brooklab/Mada-Bat-AstV.

ETHICS STATEMENT

The animal study was reviewed and approved by UC Berkeley Animal Care and Use Committee and Madagascar Ministry of Forest and the Environment under guidelines posted by the American Veterinary Medical Association.

AUTHOR CONTRIBUTIONS

CB conceived of the project and acquired the funding, in collaboration with J-MH, PD, JD, and CT. Field samples were collected and RNA extracted by AA, SA, AG, HR, TR, NR, and CB. AK led the mNGS, with support from VA, HR, TR, and CB. SH and CB analyzed the resulting data and co-wrote the original draft of the manuscript, which all authors edited and approved.

FUNDING  
Research was funded by the National Institutes of Health (1R01AI129822-01 grant to J-MH, PD, and CB), DARPA (PREEMPT Program Cooperative Agreement no. D18AC00031 to CB), the Bill and Melinda Gates Foundation (GCE/ID OPP1211841 to CB and J-MH), the Adolph C. and Mary Sprague Miller Institute for Basic Research in Science (postdoctoral fellowship to CB), the Branco Weiss Society in Science (fellowship to CB), and the Chan Zuckerberg Biohub.

ACKNOWLEDGEMENTS

The authors acknowledge Kimberly Rivera and Sarah Guth for help in the field and lab and at the Virology Unit at the Institut Pasteur de Madagascar and Maira Phelps of the Chan Zuckerburg Biohuc (CZB) for logistical support. They additionally thank Angela Detweiler, Michelle Tan, and Norma Neff of the CZB genomics platform for mNGS support and thank the Brook lab at the University of Chicago for helpful contributions to the manuscript.

SUPPLEMENTARY MATERIAL

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Description** | **Scientific Name** | **Host** | **Country** | **Max Score** | **Total Score** | **Query Cover** | **E-Value** | **Percent Identity** | **Accession Length** | **Accession** |
| Mamastrovirus sp. isolate Astrovirus\_PREDICT MAstV-132/AASMH clone 285121 RNA-dependent RNA polymerase gene, partial cds | Mamastrovirus sp. | Chiroptera | Tanzania | 302 | 302 | 5% | 8e-76 | 80.94% | 382 | KY054020 |
| Bat astrovirus isolate AstV/SC/L74.18/c2021 genomic sequence | Bat astrovirus | Hipposideros armiger | China | 71.3 | 71.3 | 0% | 2e-6 | 87.69% | 6708 | OQ236128 |
| Mamastrovirus 3 strain AstV3/Pig-wt/ESP/B377/2017, complete genome | Mamastrovirus 3 | Sus scrofa | Spain | 63.9 | 63.9 | 0% | 4e-04 | 100.00% | 6459 | MK962342 |
| Mamastrovirus 3 strain AstV3/Pig-wt/ESP/B333/2017, complete genome | Mamastrovirus 3 | Sus scrofa | Spain | 63.9 | 63.9 | 0% | 4e-04 | 100.00% | 6460 | MK962341 |
| Porcine astrovirus 3 strain USA/IA/7023/2017, complete genome | Porcine astrovirus | Porcine | USA | 63.9 | 63.9 | 0% | 4e-04 | 100.00% | 6461 | KY940545 |

da

Table S1: Top 5 NCBI blast results (nucleotide-nucleotide) for novel near-full length astrovirus genome (accession number OQ606244).

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Description** | **Scientific Name** | **Host** | **Country** | **Max Score** | **Total Score** | **Query Cover** | **E-Value** | **Percent Identity** | **Accession Length** | **Accession** |
| ORF1a [Bat astrovirus] | Bat astrovirus | Eidolon helvum | Cameroon | 691 | 691 | 40% | 0.0 | 43.09% | 891 | AWV67084 |
| Nonstructural protein [Raccoon dog astrovirus 3] | Raccoon dog astrovirus 3 | Raccoon dog | China | 538 | 973 | 62% | 0.0 | 52.43% | 1362 | ULF48000 |
| ORF1ab [Porcine astrovirus 3] | Porcine astrovirus 3 | Sus scrofa | Japan | 518 | 939 | 61% | 0.0 | 51.88% | 1353 | BAX00201 |
| ORF1ab [Porcine astrovirus 3] | Porcine astrovirus 3 | Sus scrofa | Japan | 517 | 938 | 61% | 0.0 | 51.88% | 1353 | BAX00204 |
| ORF1ab [Porcine astrovirus 3] | Porcine astrovirus 3 | Sus scrofa | Japan | 514 | 936 | 61% | 0.0 | 52.49% | 1353 | BAX00207 |

Table S2: Top 5 NCBI blastx results (nucleotide-amino acid) for novel near-full length astrovirus genome (accession number OQ606244).

LITERATURE CITED

1. Calisher, C. H., Childs, J. E., Field, H. E., Holmes, K. V. & Schountz, T. Bats: Important Reservoir Hosts of Emerging Viruses. *Clin. Microbiol. Rev.* **19**, 531–545 (2006).

2. Han, H.-J. *et al.* Bats as reservoirs of severe emerging infectious diseases. *Virus Res.* **205**, 1–6 (2015).

3. Guth, S. *et al.* Bats host the most virulent—but not the most dangerous—zoonotic viruses. *Proc. Natl. Acad. Sci.* **119**, e2113628119 (2022).

4. Luis, A. D. *et al.* A comparison of bats and rodents as reservoirs of zoonotic viruses: are bats special? *Proc. Biol. Sci.* **280**, 20122753 (2013).

5. Brook, C. E. & Dobson, A. P. Bats as ‘special’ reservoirs for emerging zoonotic pathogens. *Trends Microbiol.* **23**, 172–180 (2015).

6. Goodman, S. M., Schoeman, M. C., Rakotoarivelo, A. & Willows-Munro, S. How many species of Hipposideros have occurred on Madagascar since the Late Pleistocene? *Zool. J. Linn. Soc.* **177**, 428–449 (2016).

7. Teeling, E. C. *et al.* A Molecular Phylogeny for Bats Illuminates Biogeography and the Fossil Record. *Science* **307**, 580–584 (2005).

8. Lebarbenchon, C. *et al.* Astroviruses in bats, Madagascar. *Emerg. Microbes Infect.* **6**, 1–3 (2017).

9. Kettenburg, G. *et al.* Full Genome Nobecovirus Sequences From Malagasy Fruit Bats Define a Unique Evolutionary History for This Coronavirus Clade. *Front. Public Health* **10**, 786060 (2022).

10. Madera, S. *et al.* Discovery and Genomic Characterization of a Novel Henipavirus, Angavokely Virus, from Fruit Bats in Madagascar. *J. Virol.* **96**, e00921-22 (2022).

11. Jenkins, R. K. B. & Racey, P. A. Bats as bushmeat in Madagascar. *Madag. Conserv. Dev.* **3**, (2008).

12. Bosch, A., Pintó, R. M. & Guix, S. Human Astroviruses. *Clin. Microbiol. Rev.* **27**, 1048–1074 (2014).

13. De Benedictis, P., Schultz-Cherry, S., Burnham, A. & Cattoli, G. Astrovirus infections in humans and animals – Molecular biology, genetic diversity, and interspecies transmissions. *Infect. Genet. Evol.* **11**, 1529–1544 (2011).

14. Finkbeiner, S. R., Le, B.-M., Holtz, L. R., Storch, G. A. & Wang, D. Detection of Newly Described Astrovirus MLB1 in Stool Samples from Children. *Emerg. Infect. Dis.* **15**, 441–444 (2009).

15. Finkbeiner, S. R., Kirkwood, C. D. & Wang, D. Complete genome sequence of a highly divergent astrovirus isolated from a child with acute diarrhea. *Virol. J.* **5**, 117 (2008).

16. Chu, D. K. W. *et al.* Detection of novel astroviruses in urban brown rats and previously known astroviruses in humans. *J. Gen. Virol.* **91**, 2457–2462 (2010).

17. Kapoor, A. *et al.* Multiple novel astrovirus species in human stool. *J. Gen. Virol.* **90**, 2965–2972 (2009).

18. Fischer, K., Pinho dos Reis, V. & Balkema-Buschmann, A. Bat Astroviruses: Towards Understanding the Transmission Dynamics of a Neglected Virus Family. *Viruses* **9**, 34 (2017).

19. Karlsson, E. A. *et al.* Non-Human Primates Harbor Diverse Mammalian and Avian Astroviruses Including Those Associated with Human Infections. *PLOS Pathog.* **11**, e1005225 (2015).

20. Nagai, M. *et al.* Full genome analysis of bovine astrovirus from fecal samples of cattle in Japan: identification of possible interspecies transmission of bovine astrovirus. *Arch. Virol.* **160**, 2491–2501 (2015).

21. Ulloa, J. C. & Gutiérrez, M. F. Genomic analysis of two ORF2 segments of new porcine astrovirus isolates and their close relationship with human astroviruses. *Can. J. Microbiol.* **56**, 569–577 (2010).

22. Wolfaardt, M., Kiulia, N. M., Mwenda, J. M. & Taylor, M. B. Evidence of a recombinant wild-type human astrovirus strain from a Kenyan child with gastroenteritis. *J. Clin. Microbiol.* **49**, 728–731 (2011).

23. Prendergast, B. J., Freeman, D. A., Zucker, I. & Nelson, R. J. Periodic arousal from hibernation is necessary for initiation of immune responses in ground squirrels. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **282**, R1054-1062 (2002).

24. Kalantar, K. L. *et al.* IDseq—An open source cloud-based pipeline and analysis service for metagenomic pathogen detection and monitoring. *GigaScience* **9**, giaa111 (2020).

25. Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* **20**, 1160–1166 (2019).

26. Darriba, D. *et al.* ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. *Mol. Biol. Evol.* **37**, 291–294 (2020).

27. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference | Bioinformatics | Oxford Academic. https://academic.oup.com/bioinformatics/article/35/21/4453/5487384.

28. Posada, D. & Crandall, K. A. Felsenstein Phylogenetic Likelihood. *J. Mol. Evol.* **89**, 134–145 (2021).

29. Bouckaert, R. *et al.* BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Comput. Biol.* **15**, e1006650 (2019).

30. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data - Yu - 2017 - Methods in Ecology and Evolution - Wiley Online Library. https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.12628.

31. Fischer, K. *et al.* Insectivorous bats carry host specific astroviruses and coronaviruses across different regions in Germany. *Infect. Genet. Evol.* **37**, 108–116 (2016).

32. Lei, M. & Dong, D. Phylogenomic analyses of bat subordinal relationships based on transcriptome data. *Sci. Rep.* **6**, 27726 (2016).

33. Joffrin, L. *et al.* Astrovirus in Reunion Free-Tailed Bat (Mormopterus francoismoutoui). *Viruses* **13**, 1524 (2021).

34. Rougeron, V. *et al.* Characterization and phylogenetic analysis of new bat astroviruses detected in Gabon, Central Africa. *Acta Virol.* **60**, 386–392 (2016).

35. Drexler, J. F. *et al.* Amplification of Emerging Viruses in a Bat Colony. *Emerg. Infect. Dis.* **17**, 449–456 (2011).

36. Agnarsson, I., Zambrana-Torrelio, C. M., Flores-Saldana, N. P. & May-Collado, L. J. A time-calibrated species-level phylogeny of bats (Chiroptera, Mammalia). *PLoS Curr.* **3**, RRN1212 (2011).

37. Hoarau, F. *et al.* Bat Astrovirus in Mozambique. *Virol. J.* **15**, 104 (2018).

38. Chu, D. K. W., Poon, L. L. M., Guan, Y. & Peiris, J. S. M. Novel Astroviruses in Insectivorous Bats. *J. Virol.* **82**, 9107–9114 (2008).

39. Kemenesi, G. *et al.* Molecular Survey of RNA Viruses in Hungarian Bats: Discovering Novel Astroviruses, Coronaviruses, and Caliciviruses. *Vector-Borne Zoonotic Dis.* **14**, 846–855 (2014).

40. Lacroix, A. *et al.* Diversity of bat astroviruses in Lao PDR and Cambodia. *Infect. Genet. Evol.* **47**, 41–50 (2017).

41. Hoarau, A. O. G. *et al.* Investigation of astrovirus, coronavirus and paramyxovirus co-infections in bats in the western Indian Ocean. *Virol. J.* **18**, 205 (2021).

42. Andrianaivoarivelo, A. R., Ramilijaona, O., Racey, P., Razafindrakoto, N. & Jenkins, R. Feeding ecology, habitat use and reproduction of Rousettus madagascariensis Grandidier, 1928 (Chiroptera: Pteropodidae) in eastern Madagascar. *mammalia* **75**, (2011).