



HAL
open science

Discovery and Genomic Characterization of a Novel Henipavirus, Angavokely Virus, from Fruit Bats in Madagascar

Sharline Madera, Amy Kistler, Hafaliana Ranaivoson, Vida Ahyong, Angelo
Andrianiaina, Santino Andry, Vololoniaina Raharinosy, Tsiry H.
Randriambolamanantsoa, Ny Anjara Fifi Ravelomanantsoa, Cristina M. Tato,
et al.

► **To cite this version:**

Sharline Madera, Amy Kistler, Hafaliana Ranaivoson, Vida Ahyong, Angelo Andrianiaina, et al..
Discovery and Genomic Characterization of a Novel Henipavirus, Angavokely Virus, from Fruit Bats
in Madagascar. *Journal of Virology*, 2022, pp.e00921-22. 10.1128/jvi.00921-22 . hal-03765130

HAL Id: hal-03765130

<https://hal.science/hal-03765130v1>

Submitted on 30 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

1 Title Discovery and Genomic Characterization of a Novel Henipavirus, Angavokely virus, from
2 fruit bats in Madagascar.

3

4 Running Title Novel Henipavirus, AngV, found in Madagascar fruit bat

5

6 Authors

7 Sharline Madera¹, Amy Kistler², Hafaliana C. Ranaivoson^{3,4,5}, Vida Ahyong², Angelo

8 Andrianiaina⁵, Santino Andry⁶, Vololoniaina Raharinosy⁴, Tsiry H. Randriambolamanantsoa⁴, Ny

9 Anjara Fifi Ravelomanantsoa⁵, Cristina M. Tato², Joseph L. DeRisi^{2,7}, Hector C. Aguilar⁸, Vincent

10 Lacoste⁴, Philippe Dussart⁴, Jean-Michel Heraud^{4,9}, Cara E. Brook³

11

12 Affiliations

13 ¹ Division of Infectious Diseases, Department of Medicine, University of California, San

14 Francisco, CA, USA

15 ² Chan Zuckerberg Biohub, San Francisco, CA, USA

16 ³ Department of Ecology and Evolution, University of Chicago, Chicago, IL, USA

17 ⁴ Virology Unit, Institut Pasteur de Madagascar, Antananarivo, Madagascar

18 ⁵ Department of Zoology and Animal Biodiversity, University of Antananarivo, Antananarivo,

19 Madagascar

20 ⁶ Department of Entomology, University of Antananarivo, Antananarivo, Madagascar

21 ⁷ Department of Medicine and Biochemistry & Biophysics, University of California, San

22 Francisco, CA, USA

23 ⁸ Department of Microbiology and Immunology, College of Veterinary Medicine, Cornell
24 University, Ithaca, NY, USA

25 ⁹ Virology Department, Institut Pasteur de Dakar, Dakar, Senegal

26

27 Keywords emerging zoonosis, Henipavirus, novel virus, *Eidolon dupreanum*, bat-borne virus,

28 Madagascar

29

30 Research area Virology, Microbial genetics

31

32 Word count: 4,297

33 Abstract word count: 242 words

34 Importance count: 150 words

35 Abstract

36 The genus *Henipavirus* (family *Paramyxoviridae*) is currently comprised of seven viruses, four of
37 which have demonstrated prior evidence of zoonotic capacity. These include the biosafety level
38 4 agents Hendra (HeV) and Nipah (NiV) viruses, which circulate naturally in pteropodid fruit
39 bats. Here, we describe and characterize Angavokely virus (AngV), a divergent henipavirus
40 identified in urine samples from wild, Madagascar fruit bats. We report the near-complete
41 16,740 nt genome of AngV, which encodes the six major henipavirus structural proteins
42 (nucleocapsid, phosphoprotein, matrix, fusion, glycoprotein, and L polymerase). Within the
43 phosphoprotein (P) gene, we identify an alternative start codon encoding the AngV C protein
44 and a putative mRNA editing site where the insertion of one or two guanine residues encodes,
45 respectively, additional V and W proteins. In other paramyxovirus systems, C, V, and W are
46 accessory proteins involved in antagonism of host immune responses during infection.
47 Phylogenetic analysis suggests that AngV is ancestral to all four previously described bat
48 henipaviruses—HeV, NiV, Cedar virus (CedV), and Ghanaian bat virus (GhV)—but evolved more
49 recently than rodent- and shrew-derived henipaviruses, Mojiang (MojV), Gamak (GAKV), and
50 Daeryong (DARV) viruses. Predictive structure-based alignments suggest that AngV is unlikely to
51 bind ephrin receptors, which mediate cell entry for all other known bat henipaviruses.
52 Identification of the AngV receptor is needed to clarify the virus's potential host range. The
53 presence of V and W proteins in the AngV genome suggest that the virus could be pathogenic
54 following zoonotic spillover.

55

56

57 Importance

58 Henipaviruses include highly pathogenic emerging zoonotic viruses, derived from bat, rodent,
59 and shrew reservoirs. Bat-borne Hendra (HeV) and Nipah (NiV) are the most well-known
60 henipaviruses, for which no effective antivirals or vaccines for humans have been described.
61 Here we report the discovery and characterization of a novel henipavirus, Angavokely virus
62 (AngV), isolated from wild fruit bats in Madagascar. Genomic characterization of AngV reveals
63 all major features associated with pathogenicity in other henipaviruses, suggesting that AngV
64 could be pathogenic following spillover to human hosts. Our work suggests that AngV is an
65 ancestral bat henipavirus which likely uses viral entry pathways distinct from those previously
66 described for HeV and NiV. In Madagascar, bats are consumed as a source of human food,
67 presenting opportunities for cross-species transmission. Characterization of novel henipaviruses
68 and documentation of their pathogenic and zoonotic potential are essential to predicting and
69 preventing the emergence of future zoonoses that cause pandemics.

70

71

72

73

74

75

76

77

78

79 Introduction

80 Henipaviruses (HNVs) belong to a genus of bat-, rodent-, and shrew-borne viruses within the
81 family *Paramyxoviridae* with demonstrated zoonotic potential. HNVs can manifest extreme
82 virulence in human hosts, as exemplified by the prototypical HNVs, Hendra virus (HeV), and
83 Nipah virus (NiV), which cause severe acute respiratory distress and/or encephalitis in humans,
84 yielding case fatality rates that can exceed 90% (1–3). This high pathogenicity and the lack of
85 approved HNV therapeutics or vaccines for humans have garnered HeV and NiV classification as
86 Biological Safety Level 4 (BSL4) agents and WHO priority diseases. Since their discovery in the
87 1990s, HeV and NiV have periodically spilled over to humans from their reservoir hosts,
88 pteropodid bats. HeV zoonosis is mediated by spillover to intermediate horse hosts, from which
89 humans acquire infection (4). NiV can spillover to humans via intermediate transmission
90 through pig hosts, or directly from bat-to-human, resulting in near-annual outbreaks of fatal
91 encephalitis in South Asia, where subsequent human-to-human transmission also occurs (2, 5–
92 7).

93

94 Novel HNVs continue to emerge from wildlife hosts and represent ongoing threats to human
95 health. Initially, the *Henipavirus* genus comprised only HeV and NiV; however, the past two
96 decades have witnessed the discovery of five new HNVs: bat-borne Cedar virus (CedV) and
97 Ghanaian bat virus (GhV), rodent-borne Mojiang virus (MojV), and shrew-borne Gamak (GAKV)
98 and Daeryong viruses (DARV) (8–11). News of an eighth putative henipavirus—and the fifth
99 recognized zoonotic HNV—came to light while this article was in press: Langya henipavirus
100 (LayV), a close relative of MojV, was recently identified in febrile patients in eastern China (12).

101 Of these novel HNVs, at least three show evidence of zoonotic potential: serological data
102 suggests prior human exposure to GhV or to an antigenically-related virus in West Africa (13),
103 while MojV was first identified following a human outbreak of severe pneumonia in Chinese
104 mine workers, all of whom died after infection (9). Additionally, LayV was first identified in part
105 with sentinel surveillance of human febrile patients with a history of animal contact, then later
106 found to circulate at high seroprevalence in shrews (12). Thus, in addition to their high potential
107 for pathogenicity, HNVs possess a broad host range that spans at least seven mammalian
108 orders, including bats (10, 14).

109

110 Cross-species viral spillover necessitates effective inter-species transmission, which first
111 requires a virus to successfully enter the cells of diverse host species. In general, HNVs use the
112 highly-conserved ephrin family of proteins, both type A and type B, as cell entry receptors (1, 8,
113 15–17). A notable exception to this pattern is MojV, which does not use ephrin proteins—or the
114 sialic acid and CD150 receptors common to non-HNV paramyxoviruses—to gain cell entry (15,
115 18). Indeed, as of yet, the viral entry receptor for MojV—and the closely-related GAKV, DARV,
116 and LayV—remain unknown. In general, viruses in the genus *Henipavirus* have broad host
117 ranges and cause high case fatality rates following human spillover, making the characterization
118 of new HNVs a high public health priority.

119

120 The HNV genome consists of six structural proteins: nucleocapsid (N), phosphoprotein (P),
121 matrix (M), fusion (F), glycoprotein (G), and polymerase (L). In comparison with other members
122 of the family *Paramyxoviridae*, HNVs have relatively larger genomes (approximately 18kb vs

123 16kb). This extended length is largely due to several, long 3' untranslated regions (UTR) of the
124 N, P, F and G mRNAs (19, 20). The genome length of HNVs, like all paramyxoviruses, adheres to
125 the so-called 'Rule of Six', whereby viral genomes consistently demonstrate polyhexameric
126 length (21). The 'Rule of Six' is believed to be a requirement for efficient genome replication
127 under the unique mRNA editing features of the paramyxovirus genome (21). The paramyxovirus
128 P locus exhibits notable transcription properties that are shared across most members of the
129 *Paramyxoviridae* family. The P gene permits the translation of additional accessory proteins
130 from either gene editing events within the locus (prior to translation) or an overlapping ORF in
131 the P gene itself. All HNVs, with the exception of CedV, harbor a highly conserved mRNA editing
132 site at which the insertion of additional guanine residues can result in the translation of
133 accessory proteins, V and W, involved in viral antagonism and evasion of the host immune
134 system (1). The HNV P gene also contains an overlapping ORF that allows for the synthesis of a
135 third accessory protein, C, which is also involved in viral host immune evasion (1).

136

137 Our lab has previously presented evidence of exposure to henipa-like viruses in serum collected
138 from three endemic Madagascar fruit bat species (*E. dupreanum*, *Pteropus rufus*, and *Rousettus*
139 *madagascariensis*) using a Luminex serological assay which identified cross reactivity to
140 CedV/NiV/HeV-G and -F proteins (22). The most significant antibody binding previously
141 detected corresponded to the NiV-G antigen for *E. dupreanum* serum and the HeV-F antigen for
142 *P. rufus* and *R. madagascariensis* serum, suggesting the potential circulation of multiple HNVs in
143 the Madagascar fruit bat system (22). Fruit bats, including *E. dupreanum*, are consumed widely
144 in Madagascar as a source of human food, presenting opportunities for cross-species zoonotic

145 emergence. This underscores the importance of further characterization of the pathogenic and
146 zoonotic potential of AngV and other potential HNVs circulating in the Madagascar fruit bat
147 system. Here, we describe and characterize a novel bat HNV genome, Angavokely virus (AngV),
148 recovered from urine samples collected from the Madagascar fruit bat, *E. dupreanum*. Our
149 work suggests AngV is part of an ancestral group of HNVs that may rely on a novel, non-ephrin-
150 mediated viral entry pathway.

151

152 Methods

153 *Ethics Statement*

154 Animal capture and handling and subsequent collection of biological samples were conducted
155 in strict accordance with the Madagascar Ministry of Forest and the Environment (permit
156 numbers 019/18, 170/18, 007/19) and guidelines posted by the American Veterinary Medical
157 Association. Field protocols were approved by the UC Berkeley Animal Care and Use Committee
158 (ACUC Protocol # AUP-2017-10-10393), as previously described (23).

159

160 *Animal capture, sample collection, and RNA extraction*

161 Fruit bats were captured and processed in part with a long-term study investigating the
162 seasonal dynamics of potentially zoonotic viruses in Madagascar, as has been previously
163 described (22–26). Animals were identified morphologically by species, sex, and age class
164 (juvenile vs. adult), and urine swabs were collected into viral transport medium from any
165 individual that urinated during handling. Urine swabs were flash-frozen in liquid nitrogen in the
166 field and delivered to -80°C freezers at Institut Pasteur de Madagascar for long-term storage.

167 Urine specimens from 206 bats were randomly selected for total RNA extraction using the Zymo
168 Quick RNA/DNA Microprep Plus kit, performed as previously described (23).

169

170 *mNGS library preparation*

171 Total urine RNAs were diluted with nuclease-free H₂O, and 5uL of each specimen was used as
172 input for mNGS library preparation. A 2-fold dilution series of a 25ng/uL stock of HeLa total RNA
173 (n=8 samples), along with 5 water samples were included and processed in parallel as positive
174 and negative controls, respectively. Additionally, a 25pg aliquot of External RNA Control
175 Consortium (ERCC) spike-in mix (Thermo-Fisher) was included in each sample. Dual-indexed
176 mNGS library preparations for the samples were miniaturized and performed in 384-well
177 format with NEBNext Ultra II RNAseq library preparation kit (New England Biolabs) reagents.
178 RNA samples were fragmented for 12 min at 94°C, and 16 cycles of PCR amplification were
179 performed. Per sample read yields from a small scale iSeq (Illumina) paired-end 2 x 146bp
180 sequencing run on an equivolume pool of the individual libraries were used to normalize
181 volumes of the individual mNGS libraries to generate an equimolar pool. Paired-end 2 x 146bp
182 sequencing of the resulting equimolar library pool was performed on the NovaSeq6000
183 (Illumina) to obtain approximately 50 million reads per sample.

184

185 *Sequence analysis*

186 Raw reads from urine sample sequencing were first uploaded to the CZBID (v6.8) platform for
187 host and quality filtering and *de novo* assembly (27). In brief, in the CZBID pipeline, adaptor
188 sequences were removed with Trimmomatic (v.0.38), and reads were quality filtered (28).

189 Reads then underwent host-filtration against the Malagasy fruit bat genome, *E. dupreanum*,
190 using STAR (v 2.7.9a) (29) and a second host-removal step using Bowtie2 (30). After host
191 filtering, reads were aligned using rapsearch2 (31) and Gsnap (32), and putative pathogen taxa
192 were identified. Next, reads were assembled using SPADES (v.3.15.3) (33), and all contigs
193 generated were subject to BLAST analysis against the putative taxa previously identified by
194 rapsearch2 and Gsnap. We considered samples positive for HNV if the CZBID pipeline produced
195 at least one contig with an average read depth of two or more, which yielded a BLAST
196 alignment length >100 nt/aa and an e-value < 0.00001 (BLASTn v2.5.0+) or a bit score >100
197 (BLASTx v2.5.0+) when queried against an HNV database derived from all HNV genomes
198 available in NCBI (Accessed July 2021).

199

200 *Genome Annotation and Comparison*

201 One urine sample, collected from an adult female *E. dupreanum* fruit bat in March 2019,
202 yielded a near full-length HNV genome, which we analyzed in greater depth in subsequent
203 analyses and annotated as the novel HNV, AngV (Genbank Accession #: ON613535). Nucleotide
204 BLAST of the AngV genome identified NiV (GenBank Accession #: AF212302) as the top hit for
205 this novel virus and was subsequently chosen as the reference genome for further analysis. We
206 aligned AngV to NiV (GenBank Accession #: AF212302) in the program Geneious Prime
207 (v2020.2.4) and annotated all six major HNV structural genes, and the accessory C ORF, within
208 the P gene. We identified the putative mRNA editing site within the P gene sequence (spanning
209 nucleotides 1,225-1,232 of the P gene) and manually added one or two guanine (G) residues to
210 the 3' end of the conserved HNV mRNA editing site to generate V and W ORFs, respectively,

211 and their corresponding proteins. We furthered queried all identified transcriptional elements
212 against publicly available sequences using NCBI BLAST and BLASTx (34). Resulting BLAST and
213 BLASTx hits were used in phylogenetic analyses as described below.

214

215 We used the program pySimPlot to scan the whole genome sequence of AngV for nucleotide
216 sequence identity to the NiV genome (GenBank Accession #: AF212302) and the nucleotide and
217 amino acid sequences of individual Open Reading Frames (ORFs) contained therein.

218 Respectively, window size and scanning were specified as 50 and one for nucleotide pairwise
219 identity and 50 and five for amino acid pairwise identity. Results were visualized using Prism
220 (9.2.0).

221

222 *Phylogenetic analyses*

223 We constructed 10 Maximum Likelihood (ML) amino acid phylogenies to analyze the
224 evolutionary relatedness of our putative HNV to previously described paramyxoviruses. We
225 note that the sequence for the newly-described LayV was not yet available at the time of this
226 writing and is, therefore, absent from all phylogenies. Our (a) first ML phylogeny compared the
227 translated L protein of AngV to all translated reference L protein paramyxovirus sequences in
228 NCBI, in addition to those of the newly-described shrew HNVs, GAKV and DARV (accessed
229 November 2021). The (b) nine subsequent ML phylogenies compared the translation of each
230 individual protein annotated for AngV (N, P, C, V, W, M, F, G, L) against the top 50 BLASTx
231 sequence hits for each protein collapsed on 98% similarity. Distinct outgroups were applied: (a)
232 Sunshine Coast Virus (GenBank Accession #: YP_009094051.1) for the first L protein phylogeny,

233 (b) Human orthopneumovirus (HRSV, GenBank Accession #: NC_001781) for N, P, M, F, G, L
234 proteins, and (c) Sendai virus (GenBank Accession #: NP_056872) for gene C.

235 For each phylogenetic tree, we aligned translated amino acid sequences via the MUSCLE
236 algorithm (v3.8.1551) (35) and determined the best fit amino acid substitution model using
237 ModelTest-NG (36). Phylogenies were then constructed in RAxML-NG (37), using the
238 corresponding best fit model: JTT (complete L-protein sequence) or LG+G4+F (individual
239 proteins). In accordance with best practices outlined in the RAxML-NG manual, twenty ML
240 inferences were made on each original alignment. Bootstrap replicate trees were inferred using
241 Felsenstein's method (38). MRE-based bootstopping test was applied after every 200 replicates
242 (39), and bootstrapping was terminated once diagnostic statistics dropped below the threshold
243 value. Bootstrap support values were drawn on the best-scoring tree.

244 We additionally computed one Bayesian time-resolved phylogeny, using all 77 full-length HNV
245 nucleotide sequences available in NCBI, including our newly contributed AngV (GenBank
246 Accession #: ON613535). As with ML trees, sequences were first aligned in MUSCLE (v3.8.1551)
247 (35), and the best fit nucleotide substitution model was subsequently queried in ModelTest-NG
248 (36). We then constructed a Bayesian timetree in the program BEAST2 (40, 41), using the best
249 fit GTR+I+G4 model inferred for the whole genome alignment from ModelTest-NG and
250 assuming a constant population prior. Sampling dates corresponded to collection data as
251 reported in NCBI Virus; we assumed a collection date of 31-July in cases where only year of
252 collection was reported. We computed trees using both an uncorrelated exponentially
253 distributed relaxed molecular clock (UCED) and a strict clock but here report results from the

254 strict clock only as similar results were inferred from both. We ran Markov Chain Monte Carlo
255 (MCMC) sample chains for 1 billion iterations, checked convergence using TRACER v1.7 (42) and
256 averaged trees after 10% burn-in using TreeAnnotator v2.6.3 (43) to visualize mean posterior
257 densities at each node. The resulting phylogeny was visualized in R v.4.0.3 for MacIntosh in the
258 'ggtree' package (44).

259

260 *AngV G Protein Structure Modeling*

261 We used the Artificial Intelligence system, AlphaFold, to predict the 3D structure of the AngV
262 glycoprotein (G) (45). Molecular graphics and analyses of the AngV glycoprotein structure were
263 performed with UCSF ChimeraX (46). HNV glycoprotein ephrin binding residues were aligned
264 using the program Geneious Prime (v2020.2.4).

265

266 Results

267 *Discovery and prevalence of HNV in Malagasy bats*

268 Urine swab specimens from 206 bats were collected in 8 roosting sites across the island of
269 Madagascar from 2013 to 2019 (Figure 1). Urine samples were collected during wet and dry
270 seasons from all three Madagascar fruit bat species, *P. rufus*, *E. dupreanum*, and *R.*
271 *madagascariensis* (Table 1). Isolated RNA from urine swab specimens generated an average of
272 19 million paired-read sequences. In total, 10/206 (4.9%) bats were positive for HNV; all
273 positive samples were collected from *E. dupreanum* bats (10/106; 9.4%) at the Angavokely cave
274 roosting site (Table 1). Positive samples were collected in wet and dry seasons from both male
275 and female adults.

276

277 *Genomic characterization of AngV*

278 We recovered one near-full-length HNV contig (16,740 nt), supported by an average sequencing
279 depth of 14 reads (Figure 2A), from a urine sample collected from an adult, non-lactating *E.*
280 *dupreanum* female in the 2018-2019 wet season (capture date: 15-March 2019). We focused
281 subsequent genomic analyses on this longest sequence, which we named Angavokely virus
282 (AngV) after the site of *E. dupreanum* capture.

283

284 As with other members of the *Henipavirus* genus, the genome of AngV is organized into 6 open
285 reading frames (ORF) arranged in the order 3'-N-P-M-F-G-L-5'. AngV shares an average
286 nucleotide identity of 36% with the NiV reference genome (AF212302) and a varying amino acid
287 identity that is highest across the ORFs encoding for the nucleocapsid and L polymerase
288 proteins (Figures 2B, 2C).

289

290 *AngV coding regions*

291 The P gene of AngV follows an organization similar to most members of the *Henipavirus* genus.
292 AngV harbors alternative start sites which, respectively, encode the P and C proteins, as well as
293 a conserved putative mRNA editing site common to most paramyxoviruses A⁴⁻⁶G²⁻³ (Figure 3A
294 and B). AngV shares an identical putative mRNA editing site with the recently discovered HNVs,
295 MojV and GAKV. Pseudotemplated addition of one or two G residues at the conserved putative
296 mRNA editing site generates a putative V and W protein, respectively (Figure 3C; W protein in
297 Supplemental Figure 1). In congruence with members of the *Henipavirus* genus that encode the

298 conserved putative mRNA editing site, the putative V protein of AngV harbors a unique C-
299 terminal region that contains a highly conserved cysteine-rich zinc finger domain (Figure 3C).

300

301 The length of each ORF in the AngV genome resembles those from previously described HNVs,
302 with the exception of the gene encoding the glycoprotein (G)—which, at 688 aa, is 56 aa longer
303 than the longest previously-characterized HNV glycoprotein from GhV (632 aa; Table 2) (16).

304 BLAST analysis indicates that AngV ORFs for genes encoding the nucleocapsid (N), matrix (M),
305 and polymerase (L) proteins exhibit the highest nucleotide and amino acid pairwise identity

306 with other HNVs, with highest similarity shared with the NiV L protein (nt 74.7%, aa 52.2%,

307 Table 2). In contrast to many emerging viruses, AngV largely exhibits higher nucleotide vs.

308 amino acid identity with other HNVs (Table 2). The more recently discovered HNVs (MojV,

309 CedV, GAKV, DARV, GhV) mirror this pattern, showing higher nucleotide vs. amino acid identity

310 when compared to NiV and HeV (data not shown).

311

312 *AngV non-coding regions*

313 Examination of all viral intergenic regions (in cRNA orientation) reveals that AngV exhibits the

314 highly conserved CTT intergenic junction site characteristic of other HNVs, as well as gene stop

315 and gene start sites with high similarity to those of previously described HNVs (Supplemental

316 Table 1). We were unable to locate the intergenic junction site and transcriptional start or stop

317 site in the 5' region of the N ORF for AngV, suggesting that the genomic 3' untranslated regions

318 (UTR) for AngV have not yet been fully recovered. Comparison of the 5' and 3' UTRs for AngV

319 with those of other HNVs reveals UTRs of varying lengths within the *Henipavirus* genus

320 (Supplemental Table 2). Nevertheless, AngV exhibits similar lengths and a nucleotide identity of
321 roughly 30-40% for the 5' and 3' UTRs of most HNV genes; however, the P gene 3' UTR, the M
322 gene 5' and 3' UTRs, and the F gene 5' and 3' UTRs, are significantly shorter in AngV compared
323 with previously described HNVs. Correspondingly, nucleotide identity varies when comparing
324 this shorter subset of 5' and 3' UTRs for AngV against other HNVs (Supplemental Table 2).

325

326 *Phylogenetic Analyses*

327 Phylogenetic analysis of complete L protein amino acid sequences across the *Paramyxoviridae*
328 family places AngV within the *Henipavirus* genus at <0.82 nucleotide substitutions away from
329 the node distinguishing the family *Paramyxoviridae* from the *Sunviridae* (Figure 4A). AngV
330 clusters independently within the *Henipavirus* genus and diverges ancestral to all currently
331 known bat-borne HNVs. Our time-resolved Bayesian phylogeny further corroborates this result,
332 placing AngV ancestral to all previously described bat-borne HNVs but more recently diverged
333 than the rodent- and shrew-borne HNVs, MojV, GAKV, and DARV (Figure 4B; Supplemental
334 Figure 2). We estimate the divergence of the AngV lineage from the rest of the HNV clade at
335 9,794 years ago (95% HPD 6,519-14,024 years), and the time to the Most Recent Common
336 Ancestor (MRCA) for the entire HNV genus as 11,195 years ago (95% HPD 7,351-15,905 years).

337

338 In N, P, C, V, W, M and G amino acid phylogenies, the AngV proteins cluster closely with those
339 of other HNVs (Supplemental Figure 3). Interestingly, in the amino acid phylogeny, the AngV F
340 protein, like the F proteins of MojV, GAKV, and DARV, localizes ancestral to non-HNV
341 paramyxoviruses and distinct from the bat-borne HNV clade (Supplemental Figure 3). The AngV

342 L protein shows the highest amino acid identity to the L protein of rodent-borne Mount Mabu
343 Lophuromys virus 2 (MMLV-2), a putative Jeilongvirus (47), but is nonetheless nested between
344 the MojV/GAKV/DARV clade and the bat-borne HNV clade (Supplementary Figure 3).

345

346 *AngV glycoprotein*

347 We further examined the AngV G protein for conserved structural features and amino acid
348 residues historically associated with HNV ephrin binding. AlphaFold analysis predicted a six-
349 bladed β -propeller fold that is characteristic of *Paramyxoviridae* glycoproteins, with each blade
350 largely composed of 4 antiparallel β -strands (Figure 5A). The β -propeller fold is stabilized by
351 seven disulfide bonds that are conserved among HNVs (Supplemental Figure 4). This HNV
352 protein G structure-based alignment reveals that the elongated AngV G protein primarily
353 results from a lengthy C-terminal tail with an additional 67 aa beyond that of NiV G protein
354 (Supplemental Figure 4). Similar to the MojV G protein, the AngV G protein lacks previously
355 described ephrin binding residues (NiV W504, E505, Q530, T531, A532, E533, N557, and Y581)
356 (Figure 5B, 5C, and Supplemental Figure 4) (17, 48, 49).

357

358 Discussion

359 We describe and characterize a novel HNV, AngV, from a urine sample collected from an *E.*
360 *dupreanum* Malagasy fruit bat. In this study, urine samples from 206 unique fruit bats were
361 assessed by metagenomic sequencing, yielding an overall positive HNV detection rate of 4.9%
362 (10/206) for all bats studied and a HNV prevalence of 9.4% (10/106) for the *E. dupreanum*

363 hosts. Of all the HNV positive samples, only one sample yielded sufficient reads for assembly of
364 a complete coding sequence and subsequent genomic analysis. In a 6-year collection period
365 spanning multiple wet/dry seasons, HNV positive samples were only recovered from *E.*
366 *dupreanum* bats in the Angavokely roosting site, despite prior serological evidence of HNV
367 infection in *P. rufus* and *R. madagascariensis* bats, as well (22). HNV RNA was recovered from *E.*
368 *dupreanum* in both wet and dry seasons, though higher sampling intensity throughout the wet
369 season precludes any conclusions regarding underlying seasonal patterns in these data.
370 Previous work in this system has suggested a seasonal increase in fruit bat seroprevalence
371 across the winter low nutrient season, which also overlaps the gestation period for these
372 synchronously breeding fruit bats (22). In fruit bat systems elsewhere, HNVs are also shed in
373 urine at higher rates during the nutrient-poor dry seasons for the localities in question (50–52);
374 in the case of NiV and HeV, these seasonal viral shedding pulses have been linked to zoonotic
375 spillover.

376

377 The recovered genome of AngV exhibits a structural organization characteristic of the
378 *Henipavirus* genus and a nucleotide and amino acid identity to HeV and NiV that is comparable
379 to those shared with the more distantly related HNVs, MojV, GhV and CedV. A limited quantity
380 of available original sample precluded full genome recovery for AngV (as evidenced by the lack
381 of the 5' UTR region of the N ORF), which prevented analysis of the extent to which the full
382 AngV genome may abide by the 'Rule-of-Six', observed by all other members of the
383 *Orthoparamyxovirinae* subfamily (53). Phylogenetic analyses of AngV support classification of
384 this virus as a distinct novel bat-borne *Henipavirus* (L gene amino acid distance <0.82 distance

385 for the subfamily *Orthoparamyxovirinae*), in accordance to the International Committee on
386 Taxonomy of Viruses (ICTV) criteria (20). This novel HNV is estimated to have diverged
387 approximately 9,800 years ago, prior to the currently known African and Asian bat-borne HNV
388 lineages but considerably more recently than the estimated mid-to late-Miocene divergence of
389 *E. dupreanum* from its sister species, *E. helvum*, on the African continent (54). Recent
390 characterization of *Betacoronaviruses* in Madagascar fruit bats demonstrates surprising identity
391 to lineages circulating in West Africa (23), suggesting that, though resident only in Madagascar,
392 Malagasy fruit bats likely experience some form of contact with the African continent. Of the 49
393 bat species that inhabit the island nation of Madagascar, nine species are widely distributed
394 across Africa, Asia, and/or Europe, presenting opportunities for inter-species viral transmission
395 via island-hopping. Intensified viral sampling of Madagascar's insectivorous bat populations for
396 HNVs thus represents an important future research priority.

397

398 As an ancestral bat-borne HNV, AngV may provide important insight into HNV evolution and
399 pathogenesis. Similar to other paramyxoviruses, the encoded AngV P gene is able to produce
400 multiple immunomodulatory protein products (55). One such protein product is the V protein,
401 thought to be involved in immune evasion and considered a significant determinant of viral
402 pathogenicity and lethality (56, 57). AngV harbors the highly conserved mRNA editing site and a
403 predicted ORF that encodes a V protein with a conserved cysteine-rich C-terminus, suggesting
404 that AngV has the capacity to produce a functional V protein. With the exception of the newly
405 discovered HNVs in shrews, GAKV and DARV, all HNVs harboring a V protein have previously
406 demonstrated evidence of human infection, highlighting the potential for AngV to cause

407 productive infection in humans (1, 9, 13). Further studies are needed to ascertain the virulence
408 potential and host breadth of this novel virus.

409

410 Characterization of the AngV glycoprotein (G) through AlphaFold modeling and structure-based
411 alignments revealed a similar structural organization to other HNV glycoproteins. Notably, the
412 AngV glycoprotein surpasses that of GhV as the longest glycoprotein of the *Henipavirus* genus.
413 Like that of GhV, the AngV glycoprotein harbors a long C terminal extension (Supplemental
414 Figure 4). It is unclear if the C terminal extension of the AngV glycoprotein has a functional role,
415 though the C terminal extension of the glycoprotein in GhV is known to play a functional role in
416 receptor-mediated fusion (16).

417

418 Henipavirus host tropism and virulence rely on a myriad of factors, one of which is the HNV
419 glycoprotein. The previously characterized HNV glycoproteins of NiV, HeV, CedV, and GhV,
420 utilize members of the ephrinA and ephrinB class family as host-cell receptors for viral entry
421 into human cells (16–18, 48, 58). However, like MojV, the AngV glycoprotein lacks these well-
422 conserved ephrin binding residues. Structure-based alignments can shed light on potential
423 receptor binding residues when characterizing novel viruses. For instance, sequence-based
424 comparisons of the GhV and NiV glycoproteins were used to predict GhV ephrin binding (13),
425 which was later confirmed by crystallography (16). Structure-based alignment of the AngV
426 glycoprotein shows a lack of highly conserved ephrin binding residues, including NiV E533 – a
427 seminal residue for ephrinB2 binding that is conserved across all ephrin binding HNVs. This
428 suggests that, like MojV—and probably DARV, GAKV, and LayV—the AngV glycoprotein may not

429 bind ephrins, pointing to the possible use of an ancestral viral entry pathway. The growing
430 number of novel HNVs that appear not to rely on ephrin binding for cellular entry could warrant
431 re-evaluation of the existing HNV genus to better reflect conserved function and pathobiology.

432

433 This work presents a novel bat-HNV, AngV, identified from a Malagasy fruit bat. AngV joins a
434 growing group of ancestral HNVs with unknown cell-entry receptors. Discovery of the cell
435 surface receptor for AngV represents an important future research priority that will shed light
436 on the breadth of host range for this virus, including its zoonotic potential.

437

438

439

440

441 Acknowledgements

442 The authors thank Anecia Gentles and Kimberly Rivera for help in the field and the lab and
443 acknowledge the Virology Unit at the Institut Pasteur de Madagascar and Maira Phelps of the
444 Chan Zuckerberg Biohub (CZB) for logistical support. They additionally thank Angela Detweiler,
445 Michelle Tan, and Norma Neff of the CZB genomics platform for mNGS support. Molecular
446 graphics and analyses were performed with UCSF ChimeraX, developed by the Resource for
447 Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with
448 support from the National Institutes of Health R01-GM129325 and the Office of Cyber
449 Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases.

450

451 Funding

452 This work was supported by the National Institutes of Health (1R01AI129822-01 grant to JMH,
453 PD, and CEB; 5T32AI007641-19 to SM; R01AI109022 to HAC), DARPA (PREEMPT Program
454 Cooperative Agreement no. D18AC00031 to CEB), the Bill and Melinda Gates Foundation
455 (GCE/ID OPP1211841 to CEB and JMH), the Adolph C. and Mary Sprague Miller Institute for
456 Basic Research in Science (postdoctoral fellowship to CEB), AAAS/L'Oréal USA (For Women in
457 Science fellowship to CEB), the Branco Weiss Society in Science (fellowship to CEB), and the
458 Chan Zuckerberg Biohub.

459

460 Conflict of Interests

461 The authors declare no competing interests.

462

463 Data Availability

464 Raw and assembled sequencing data are deposited in NCBI Bioproject PRJNA837298. The full
465 genome of AngV is available in GenBank under Accession # ON613535. All raw data and code
466 for figures can be obtained in our open-access GitHub repository:

467 <https://github.com/brooklabteam/angavokely-virus>

468

469

470

471

472

- 474 1. Eaton BT, Broder CC, Middleton D, Wang L-F. 2006. Hendra and Nipah viruses:
475 Different and dangerous. *Nat Rev Microbiol* 4:23–35.
- 476 2. Sharma V, Kaushik S, Kumar R, Yadav JP, Kaushik S. 2019. Emerging trends of
477 Nipah virus: A review. *Reviews in Medical Virology* 29.
- 478 3. Arunkumar G, Chandni R, Mourya DT, Singh SK, Sadanandan R, Sudan P, Bhargava
479 B. 2019. Outbreak investigation of Nipah Virus Disease in Kerala, India, 2018.
480 *Journal of Infectious Diseases* 219:1867–1878.
- 481 4. Murray K, Selleck P, Hooper P, Hyatt a, Gould a, Gleeson L, Westbury H, Hiley L,
482 Selvey L, Rodwell B. 1995. A morbillivirus that caused fatal disease in horses and
483 humans. *Science* 268:94–7.
- 484 5. Hsu VP, Hossain MJ, Parashar UD, Ali MM, Ksiazek TG, Kuzmin I, Niezgoda M,
485 Rupprecht C, Bresee J, Breiman RF. 2004. Nipah virus encephalitis reemergence,
486 Bangladesh. *Emerging Infectious Diseases* 10:2082–2087.
- 487 6. Gurley ES, Montgomery JM, Hossain MJ, Bell M, Azad AK, Islam MR, Molla MAR,
488 Carroll DS, Ksiazek TG, Rota PA, Lowe L, Comer JA, Rollin P, Czub M, Grolla A,
489 Feldmann H, Woodward JL, Breiman RF. 2007. Person-to-Person transmission of
490 Nipah virus in a Bangladeshi community. *Emerging Infectious Diseases* 13:1031–
491 1037.
- 492 7. Luby SP, Gurley ES. 2012. Epidemiology of Henipavirus Disease in Humans, p. 25–
493 40. *In* Lee, B, Rota, PA (eds.), *Henipavirus*. Springer Berlin Heidelberg, Berlin,
494 Heidelberg.
- 495 8. Marsh GA, de Jong C, Barr JA, Tachedjian M, Smith C, Middleton D, Yu M, Todd S,
496 Foord AJ, Haring V, Payne J, Robinson R, Broz I, Crameri G, Field HE, Wang L-F.
497 2012. Cedar virus: a novel Henipavirus isolated from Australian bats. *PLoS*
498 *Pathogens* 8:e1002836.
- 499 9. Wu Z, Yang L, Yang F, Ren X, Jiang J, Dong J, Sun L, Zhu Y, Zhou H, Jin Q. 2014.
500 Novel Henipa-like virus, Mojiang paramyxovirus, in rats, China, 2012. *Emerging*
501 *Infectious Diseases* 20:1064–1066.
- 502 10. Lee SH, Kim K, Kim J, No JS, Park K, Budhathoki S, Lee SH, Lee J, Cho SH, Cho S, Lee
503 GY, Hwang J, Kim HC, Klein TA, Uhm CS, Kim WK, Song JW. 2021. Discovery and
504 genetic characterization of novel paramyxoviruses related to the genus
505 henipavirus in *Crocodylus* species in the republic of Korea. *Viruses* 13.
- 506 11. Drexler JF, Corman VM, Müller MA, Maganga GD, Vallo P, Binger T, Gloza-Rausch
507 F, Rasche A, Yordanov S, Seebens A, Oppong S, Adu Sarkodie Y, Pongombo C,
508 Lukashov AN, Schmidt-Chanasit J, Stöcker A, Carneiro AJB, Erbar S, Maisner A,
509 Fronhoffs F, Buettner R, Kalko EK v, Kruppa T, Franke CR, Kallies R, Yandoko ERN,
510 Herrler G, Reusken C, Hassanin A, Krüger DH, Matthee S, Ulrich RG, Leroy EM,
511 Drosten C. 2012. Bats host major mammalian paramyxoviruses. *Nat Commun*
512 3:796.
- 513 12. Zhang X-A, Li H, Jiang F-C, Zhu F, Zhang Y-F, Chen J-J, Tan C-W, Anderson DE, Fan H,
514 Dong L-Y, Li C, Zhang P-H, Li Y, Ding H, Fang L-Q, Wang L-F, Liu W. 2022. A zoonotic

- 515 henipavirus in febrile patientis in China. *New England Journal of Medicine*
516 387:468–470.
- 517 13. Pernet O, Schneider BS, Beaty SM, LeBreton M, Yun TE, Park A, Zachariah TT,
518 Bowden T a., Hitchens P, Ramirez CM, Daszak P, Mazet J, Freiberg AN, Wolfe ND,
519 Lee B. 2014. Evidence for henipavirus spillover into human populations in Africa.
520 *Nature Communications* 5:5342.
- 521 14. Amaya M, Broder CC. 2020. Vaccines to emerging viruses: Nipah and Hendra.
522 *Annual Review of Virology* 7:447–473.
- 523 15. Cheliout Da Silva S, Yan L, Dang H v., Xu K, Epstein JH, Veessler D, Broder CC. 2021.
524 Functional analysis of the fusion and attachment glycoproteins of mojiang
525 henipavirus. *Viruses* 13.
- 526 16. Lee B, Pernet O, Ahmed A a., Zeltina A, Beaty SM, Bowden T a. 2015. Molecular
527 recognition of human ephrinB2 cell surface receptor by an emergent African
528 henipavirus. *Proceedings of the National Academy of Sciences* 201501690.
- 529 17. Laing ED, Navaratnarajah CK, Cheliout S, Silva D, Petzing SR, Xu Y, Broder CC, Xu K.
530 2019. Structural and functional analyses reveal promiscuous and species specific
531 use of ephrin receptors by Cedar virus. *Proceedings of the National Academy of*
532 *Sciences* 116:20707–20715.
- 533 18. Rissanen I, Ahmed AA, Azarm K, Beaty S, Hong P, Nambulli S, Duprex WP, Lee B,
534 Bowden TA. 2017. Idiosyncratic Mòjiang virus attachment glycoprotein directs a
535 host-cell entry pathway distinct from genetically related henipaviruses. *Nature*
536 *Communications* 8.
- 537 19. Flick R, Walpita P, Czub M. 2006. Nipah and Hendra viral infections, p. 586–589. *In*
538 *Tropical Infectious Diseases*. Churchill Livingstone.
- 539 20. Rima B, Buschmann AB-, Dundon WG, Duprex P, Easton A, Fouchier R, Kurath G,
540 Lamb R, Lee B, Rota P, Wang L, Consortium IR. 2019. ICTV Virus Taxonomy Profile:
541 *Paramyxoviridae*. *Journal of General Virology* 100:1593–1594.
- 542 21. Calain P, Roux L. 1993. The rule of six, a basic feature for efficient replication of
543 Sendai virus defective interfering RNA. *Journal of Virology* 67:4822–4830.
- 544 22. Brook CE, Ranaivoson HC, Broder CC, Cunningham AA, Héraud J-M, Peel AJ, Gibson
545 L, Wood JLN, Metcalf CJ, Dobson AP. 2019. Disentangling serology to elucidate
546 henipa- and filovirus transmission in Madagascar fruit bats. *Journal of Animal*
547 *Ecology* 88:1001–1016.
- 548 23. Kettenburg G, Kistler A, Ranaivoson HC, Ahyong V, Andrianiana A, Andry S, DeRisi
549 JL, Gentles A, Raharinosy V, Randriambolamanantsoa TH, Ravelomanantsoa NAF,
550 Tato CM, Dussart P, Heraud J-M, Brook CE. 2022. Full genome Nobecovirus
551 sequences from Malagasy fruit bats define a unique evolutionary history for this
552 coronavirus clade. *Frontiers in Public Health* 10.
- 553 24. Brook CE, Ranaivoson HC, Andriafidison D, Ralisata M, Razafimanahaka J, Héraud J,
554 Dobson AP, Metcalf CJ. 2019. Population trends for two Malagasy fruit bats.
555 *Biological Conservation* 234:165–171.
- 556 25. Ranaivoson HC, Héraud J-M, Goethert HK, Telford III SR, Rabetafika L, Brook CE.
557 2019. Babesial infection in the Madagascar flying fox, *Pteropus rufus* É. Geoffroy,
558 1803. *Parasites & Vectors* 1–13.

- 559 26. Brook CE, Bai Y, Dobson AP, Osikowicz LM, Ranaivoson HC, Zhu Q, Kosoy MY,
560 Dittmar K. 2015. *Bartonella* spp. in Fruit Bats and Blood-Feeding Ectoparasites in
561 Madagascar. *PLoS Neglected Tropical Diseases* 9.
- 562 27. Kalantar KL, Carvalho T, de Bourcy CFA, Dimitrov B, Dingle G, Egger R, Han J,
563 Holmes OB, Juan YF, King R, Kislyuk A, Lin MF, Mariano M, Morse T, Reynoso L v.,
564 Cruz DR, Sheu J, Tang J, Wang J, Zhang MA, Zhong E, Ahyong V, Lay S, Chea S, Bohl
565 JA, Manning JE, Tato CM, DeRisi JL. 2021. IDseq-An open source cloud-based
566 pipeline and analysis service for metagenomic pathogen detection and
567 monitoring. *Gigascience* 9:1–14.
- 568 28. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina
569 sequence data. *Bioinformatics* 30:2114–2120.
- 570 29. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M,
571 Gingeras TR. 2013. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*
572 29:15–21.
- 573 30. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature*
574 9:357–360.
- 575 31. Zhao Y, Tang H, Ye Y. 2012. RAPSearch2: A fast and memory-efficient protein
576 similarity search tool for next-generation sequencing data. *Bioinformatics* 28:125–
577 126.
- 578 32. Wu TD, Nacu S. 2010. Fast and SNP-tolerant detection of complex variants and
579 splicing in short reads. *Bioinformatics* 26:873–881.
- 580 33. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM,
581 Nikolenko SI, Pham S, Prjibelski AD, Pyshkin A v., Sirotkin A v., Vyahhi N, Tesler G,
582 Alekseyev MA, Pevzner PA. 2012. SPAdes: A new genome assembly algorithm and
583 its applications to single-cell sequencing. *Journal of Computational Biology*
584 19:455–477.
- 585 34. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment
586 search tool. *Journal of Molecular Biology* 215:403–410.
- 587 35. Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and
588 high throughput. *Nucleic Acids Research* 32:1792–1797.
- 589 36. Darriba Di, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. 2020.
590 ModelTest-NG: A new and scalable tool for the selection of DNA and protein
591 evolutionary models. *Molecular Biology and Evolution* 37:291–294.
- 592 37. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAXML-NG: A fast,
593 scalable and user-friendly tool for maximum likelihood phylogenetic inference.
594 *Bioinformatics* 35:4453–4455.
- 595 38. Felsenstein J. 1985. Confidence limits on phylogenies: An approach using the
596 bootstrap. *Evolution (N Y)* 39:783–791.
- 597 39. Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A. 2010.
598 How many bootstrap replicates are necessary? *Journal of Computational Biology*
599 17:337–354.
- 600 40. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut
601 A, Drummond AJ. 2014. BEAST 2: A software platform for Bayesian evolutionary
602 analysis. *PLoS Computational Biology* 10.

- 603 41. Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with
604 BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* 29:1969–1973.
- 605 42. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior
606 summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*
607 67:901–904.
- 608 43. Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by
609 sampling trees. *BMC Evolutionary Biology* 7.
- 610 44. Yu G, Smith DK, Zhu H, Guan Y, Lam TTY. 2017. Ggtree: an R Package for
611 visualization and annotation of phylogenetic trees with their covariates and other
612 associated data. *Methods in Ecology and Evolution* 8:28–36.
- 613 45. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O,
614 Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl
615 SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T,
616 Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M,
617 Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli
618 P, Hassabis D. 2021. Highly accurate protein structure prediction with AlphaFold.
619 *Nature* 596:583–589.
- 620 46. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH,
621 Ferrin TE. 2021. UCSF ChimeraX: Structure visualization for researchers, educators,
622 and developers. *Protein Science* 30:70–82.
- 623 47. Vanmechelen B, Bletsa M, Laenen L, Lopes AR, Vergote V, Beller L, Deboutte W,
624 Korva M, Avšič Županc T, Goüy de Bellocq J, Gryseels S, Leirs H, Lemey P, Vrancken
625 B, Maes P. 2018. Discovery and genome characterization of three new
626 Jeilongviruses, a lineage of paramyxoviruses characterized by their unique
627 membrane proteins. *BMC Genomics* 19.
- 628 48. Guillaume V, Aslan H, Ainouze M, Guerbois M, Fabian Wild T, Buckland R,
629 Langedijk JPM. 2006. Evidence of a potential receptor-binding site on the Nipah
630 virus G protein (NiV-G): Identification of globular head residues with a role in
631 fusion promotion and their localization on an NiV-G structural model. *Journal of*
632 *Virology* 80:7546–7554.
- 633 49. Negrete OA, Wolf MC, Aguilar HC, Enterlein S, Wang W, Mühlberger E, Su S v.,
634 Bertolotti-Ciarlet A, Flick R, Lee B. 2006. Two key residues in EphrinB3 are critical
635 for its use as an alternative receptor for Nipah virus. *PLoS Pathogens* 2:0078–0086.
- 636 50. Field H, Jordan D, Edson D, Morris S, Melville D, Parry-Jones K, Broos A, Divljan A,
637 McMichael L, Davis R, Kung N, Kirkland P, Smith C. 2015. Spatiotemporal Aspects
638 of Hendra Virus Infection in Pteropid Bats (Flying-Foxes) in Eastern Australia. *Plos*
639 *One* 10:e0144055.
- 640 51. Páez DJ, Giles J, Mccallum H, Field H, Jordan D, Peel AJ, Plowright RK. 2017.
641 Conditions affecting the timing and magnitude of Hendra virus shedding across
642 pteropodid bat populations in Australia. *Epidemiology and Infection* 145:3143–
643 3153.
- 644 52. Cappelle J, Furey N, Hoem T, Ou TP, Lim T, Hul V, Heng O, Chevalier V, Dussart P,
645 Duong V. 2021. Longitudinal monitoring in Cambodia suggests higher circulation of

646 alpha and betacoronaviruses in juvenile and immature bats of three species.
647 Scientific Reports 11:24145.

648 53. Hausmann S, Jacques J-P, Kolakofsky D. 1996. Paramyxovirus RNA editing and the
649 requirement for hexamer genome length. RNA 2:1033–1045.

650 54. Shi JJ, Chan LM, Peel AJ, Lai R, Yoder AD, Goodman SM. 2014. A deep divergence
651 time between sister species of eidolon (Pteropodidae) with evidence for
652 widespread panmixia. Acta Chiropterologica 16:279–292.

653 55. Douglas J, Drummond AJ, Kingston RL. 2021. Evolutionary history of
654 cotranscriptional editing in the paramyxoviral phosphoprotein gene. Virus
655 Evolution 7.

656 56. Satterfield BA, Cross RW, Fenton KA, Agans KN, Basler CF, Geisbert TW, Mire CE.
657 2015. The immunomodulating v and W proteins of Nipah virus determine disease
658 course. Nature Communications 6.

659 57. Patterson JB, Thomas D, Lewicki H, Billeter MA, Oldstone MBA. 2000. V and C
660 proteins of measles virus function as virulence factors in vivo. Virology 267:80–89.

661 58. Negrete OA, Chu D, Aguilar HC, Lee B. 2007. Single amino acid changes in the
662 Nipah and Hendra virus attachment glycoproteins distinguish ephrinB2 from
663 ephrinB3 usage. Journal of Virology 81:10804–10814.

664

665

666

667

668

669

670

671

672

673

674

675

676

677 **Table 1.** Prevalence of HNV infections in the urine of Malagasy bats captured during 2013-2019
 678 collection period.

Roost site	Species	By Site		Season	By Season/Sex	
		Total sampled	Total henipavirus positive		Total sampled # (M,F)	Total henipavirus positive # (M,F)
Ambakoana	<i>P. rufus</i>	18	0	wet '13	2 (1,1)	0 (0,0)
				wet '14/'15	3 (3,0)	
				wet '15/'16	2 (2,0)	
				wet '17/'18	2 (2,0)	
				wet '18/'19	9 (6,3)	
Angavobe	<i>E. dupreanum</i>	8	0	wet '17/'18	5 (1,4)	0 (0,0)
				dry '18	3 (1,2)	0 (0,0)
Angavokely	<i>E. dupreanum</i>	98	10	wet '15/'16	2 (1,1)	0 (0,0)
				wet '17/'18	38 (5,33)	4 (0,4)
				dry '18	11 (8,3)	2 (1,1)
				wet '18/'19	47 (20,27)	4 (2,2)
Mahabo	<i>P. rufus</i>	8	0	dry '14	8 (4,4)	0 (0,0)
Mahialambo	<i>P. rufus</i>	5	0	wet '18/'19	5 (2,3)	0 (0,0)
Makira	<i>P. rufus</i>	2	0	dry '15	2 (2,0)	0 (0,0)
Marovitsika	<i>P. rufus</i>	62	0	wet '17/'18	9 (3,6)	0 (0,0)
				dry '18	7 (3,4)	
				wet '18/'19	46 (20,26)	
Maromizaha	<i>R. madagascariensis</i>	5	0	wet '13	2 (1,1)	0 (0,0)
				dry '14	1 (0,1)	
				wet '14/'15	2 (2,0)	

679

680

681

682

683

684

685

686

687 **Table 2.** Length and pairwise sequence identity of predicted open reading frames of AngV and
 688 other HNV.

689

Gene	AngV length aa	NiV length aa (%nt, %aa)	HeV	MojV	CedV	GAKV	DARV	GhV
N	514	532 (56.4, 48.0)	532 (55.4, 46.7)	539 (55.8, 45.3)	510 (57.5, 48.3)	533 (58.0, 47.5)	574 (55.3, 24.0)	514 (56.2, 48.9)
P	693	709 (47.8, 25.1)	707 (48.1, 24.1)	694 (47.6, 23.6)	737 (42.3, 21.6)	586 (43.8, 22.5)	698 (47.3, 23.2)	870 (42.2, 23.1)
C	173	166 (51.3, 25.7)	166 (51.5, 25.1)	177 (47.6, 22.5)	177 (54.5, 29.8)	184 (47.4, 27.5)	175 (48.1, 24.0)	163 (46.1, 23.1)
V	461	456 (46.9, 22.8)	457 (48.5, 23.6)	464 (47.2, 24.5)		370 (43.0, 22.3)	468 (47.2, 22.2)	621 (24.7, 19.8)
W	412	450 (47.0, 21.3)	448 (45.0, 20.0)	435 (46.5, 22.5)		331 (43.9, 22.7)	437 (45.5, 21.1)	572 (39.3, 16.8)
M	354	352 (60.0, 54.6)	352 (59.5, 54.0)	340 (57.1, 53.4)	360 (56.9, 53.1)	340 (56.1, 50.8)	345 (56.5, 52.3)	343 (56.9, 51.4)
F	539	546 (53.7, 39.4)	546 (51.7, 39.4)	545 (52.2, 40.6)	557 (51.8, 36.3)	565 (53.3, 39.4)	545 (58.8, 40.1)	662 (45.5, 35.4)
G	688	602 (43.5, 21.4)	604 (43.8, 20.4)	625 (43.0, 21.2)	622 (43.3, 19.8)	635 (44.9, 17.8)	628 (44.3, 18.7)	632 (42.5, 19.8)
L	2,259	2,244 (74.7, 52.2)	2,244 (57.6, 51.9)	2,277 (57.5, 51.8)	2,501 (52.4, 44.8)	2,291 (57.7, 51.0)	2,271 (58.8, 51.3)	2,250 (57.0, 49.1)

690

691

692

693

694

695

696

697

698

699

700

701

702 **Figure Legends**

703 **Figure 1.** Geographic location of sampling sites used in this study. Sampling sites grouped by bat
704 species found depicted as follows *P. rufus* (pink circles) Ambakoana (-18.51 S, 48.17 E) /
705 Mahabo (-20.46 S, 44.68 E) / Mahialambo (-18.11 S, 48.21 E) / Makira (-15.11 S, 49.59 E) /
706 Marovitsika (-18.84 S, 48.06 E) roosts; *E. dupreanum* (green triangles) Angavobe (-18.94 S,
707 47.95 E) /Angavokely (-18.93 S, 47.76 E) caves; *R. madagascariensis* (orange squares)
708 Maromizaha cave (-18.96 S, 48.45). Dashed lines link each sampling site to corresponding pie
709 charts, which show the percentage of HNV positive specimens for all sampled species and sites.
710 Dashed lines are included for visualization purposes only, and variation in length is not
711 significant. Pies are size-weighted by the total bat population sampled at each site,
712 corresponding to the legend. HNV positive samples were only recovered from the *E.*
713 *dupreanum* Angavokely site.

714

715 **Figure 2.** AngV genome organization. A. Coding regions for each gene are shown and depicted
716 in color, non-coding intergenic and terminal regions are highlighted in gray. Depicted genes
717 represented as follows: nucleocapsid (N), RNA polymerase (P and L), matrix (M), fusion (F), and
718 glycoprotein (G). Sequencing read depth supporting each position of the recovered genome
719 sequence is plotted below the genomic schematic. Scanning nucleotide (B) and amino acid (C)
720 pairwise identity to Nipah virus (GenBank Accession #: AF212302). Dotted horizontal lines
721 represent average read depth (14.29) or average nucleotide pairwise identity (36%).

722

723 **Figure 3.** Organization of the P gene of AngV. A. Alternative transcriptional start sites (pink
724 triangle) generate the P and C protein. Pseudotemplated addition of one or two guanine
725 nucleotides at the putative mRNA editing site generates a V and W protein, respectively. B.
726 Sequence alignment of the putative mRNA editing site across members of the *Henipavirus*
727 genus (cRNA depicted). C. Amino acid alignment of the unique C terminal region of the V
728 protein following the addition of one guanine nucleotide to the putative mRNA editing site.
729 Gray boxes denote conserved cysteine and histidine residues suggested to directly coordinate
730 bound zinc ions (55). Individual nucleotides or amino acids are color coordinated if at least 75%
731 conserved at the alignment position. Nucleotide or amino acid position numbers displayed
732 represent the position within the AngV gene or protein. Virus name (abbreviation), followed by
733 GenBank Accession #: Angavokely virus (AngV) ON613535; Nipah virus (NiV) AF212302; Hendra
734 virus (HeV) AF017149; Mojiang virus (MojV) KF278639; Ghanaian bat Henipavirus (GhV)
735 HQ660129; Daeryong virus (DARV) MZ574409; Gamak virus (GAKV) MZ574407, Cedar virus
736 (CedV) JQ001776. CedV is shown here only for comparison, as the CedV P protein is not
737 believed to undergo RNA editing or to generate a functional V protein (8, 55).

738

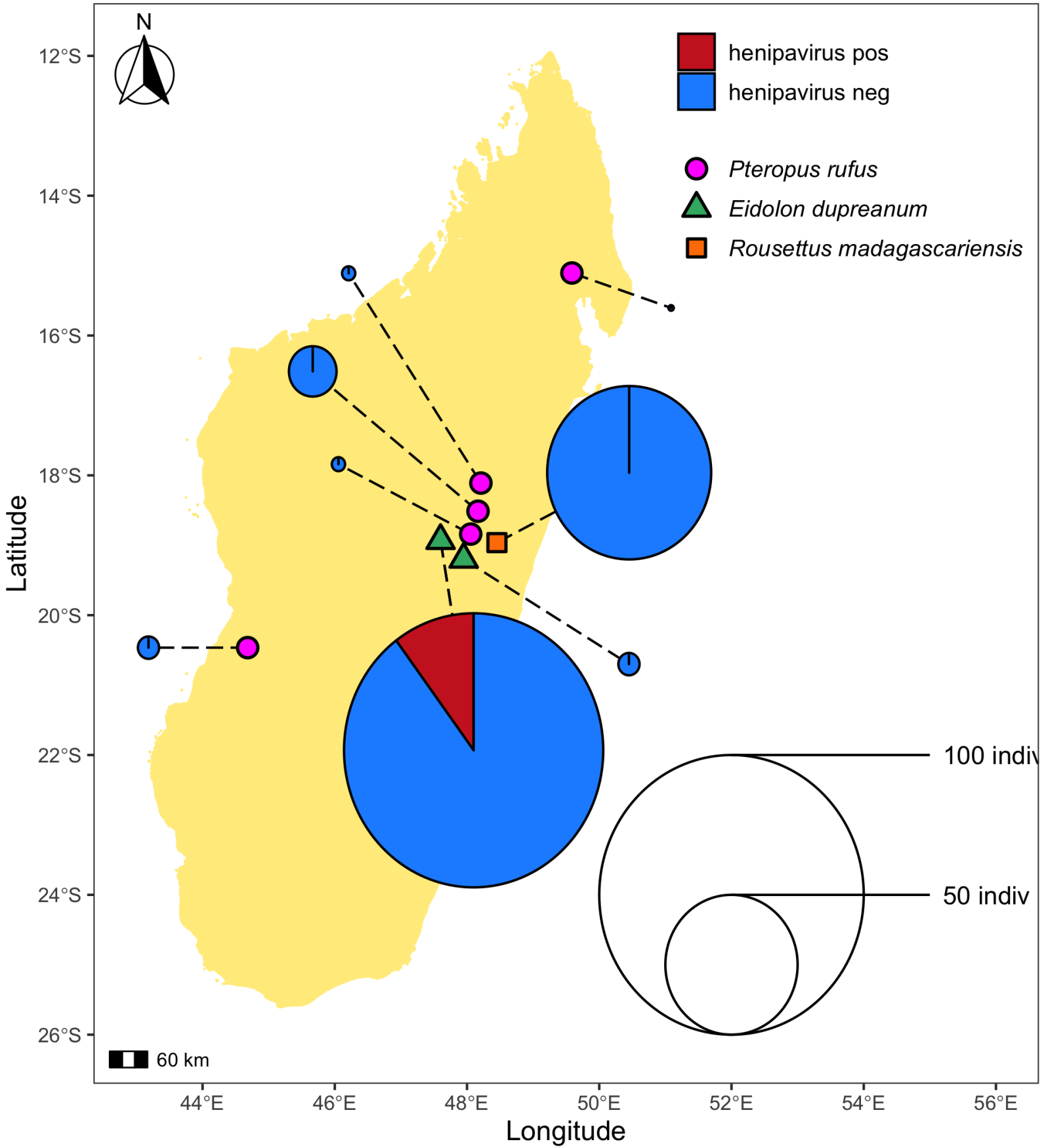
739 **Figure 4.** A. Phylogenetic analysis of the complete L protein sequences of members of the
740 family *Paramyxoviridae*. We note that the sequence for the newly-described Langya
741 henipavirus (LayV) (12) was not yet available at the time of this writing and is, therefore, not
742 included in the phylogenies. Tree is rooted with Sunshine Coast Virus (GenBank Accession #:
743 YP_009094051.1) as an outgroup, with outgroup branch length shrunk for ease of viewing.
744 Novel HNV, AngV, is depicted in green. Subfamilies and genera are demarcated, excluding those

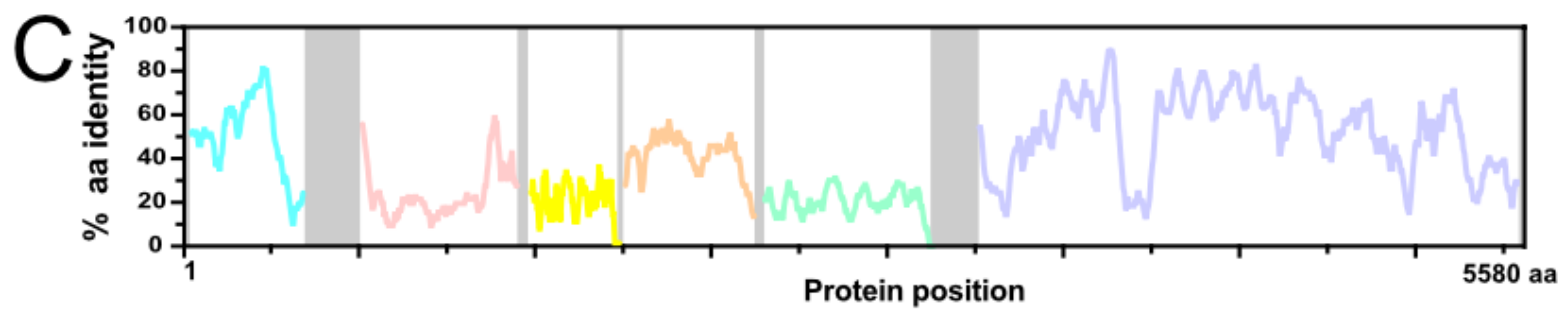
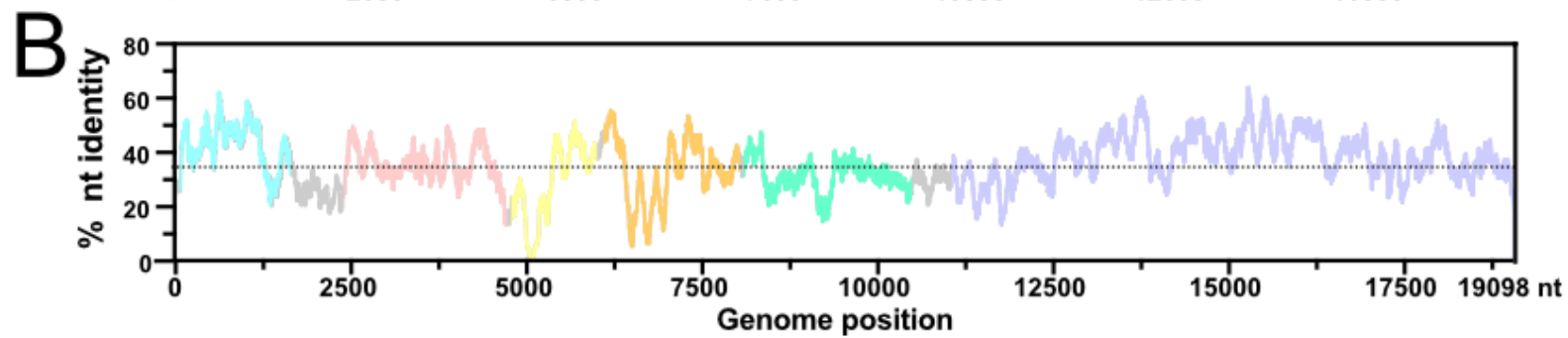
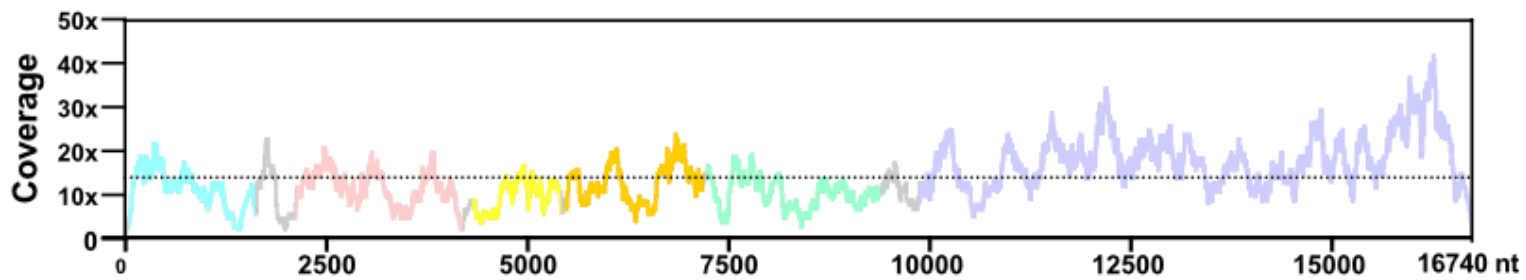
745 unassigned to subfamily (genera *Scolidonvirus*, *Cynoglossusvirus*, *Hoplichthysvirus*). Bootstrap
746 support is depicted and GenBank Accession numbers displayed next to virus names. Scale bar
747 represents substitutions per site. B. Time-resolved Bayesian phylogeny computed in BEAST2
748 incorporating all available *Henipavirus* whole genome nucleotide sequences, with the addition
749 of newly discovered GAKV, DARV, and AngV. Closely-related sequences are collapsed at triangle
750 nodes for NiV and HeV (phylogeny with un-collapsed branches available in Supplemental Figure
751 2). 95% HPD intervals around the timing of each branching node are visualized as red horizontal
752 bars. Posterior support $>.9$ is indicated by black coloring of the corresponding node, and distinct
753 *Henipavirus* species are indicated by colored tip points, with AngV highlighted in yellow for
754 further emphasis. The estimated time to MRCA for Angavokely virus and the previously-
755 described bat-borne HNVs is 9,794 (95% HPD 6,519 – 14,025) years ago.

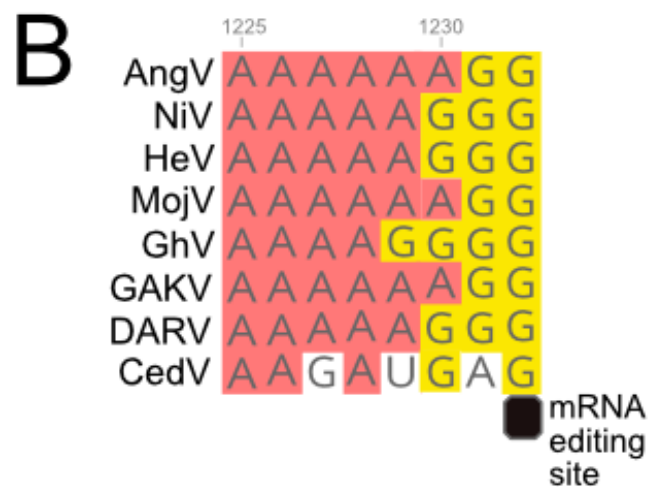
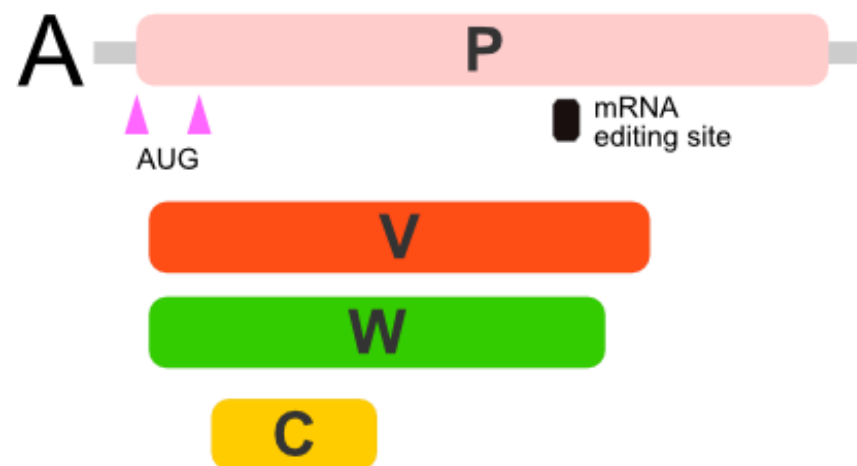
756

757 **Figure 5.** AlphaFold-predicted AngV glycoprotein 3D structure and ephrin binding residue
758 sequence alignment. A. AlphaFold-predicted 3D structure of AngV glycoprotein. N and C
759 termini are indicated in white text, and residues corresponding to ephrin binding sites in other
760 HNVs that are not conserved in AngV are colored green. B. Alignment of HNV ephrin binding
761 residues. The position of previously-described HNV ephrin binding residues are noted by a star,
762 and residues conserved across most HNVs are highlighted yellow. Amino acid position numbers
763 displayed represent the position within the AngV or NiV glycoproteins. Virus name
764 (abbreviations) followed by GenBank Accession #: Angavokely virus (AngV) ON613535; Nipah
765 virus (NiV) AF212302; Hendra virus (HeV) AF017149; Mojiang virus (MojV) KF278639; Cedar
766 virus (CedV) JQ001776; Ghanaian bat Henipavirus (GhV) HQ660129; Gamak virus (GAKV)

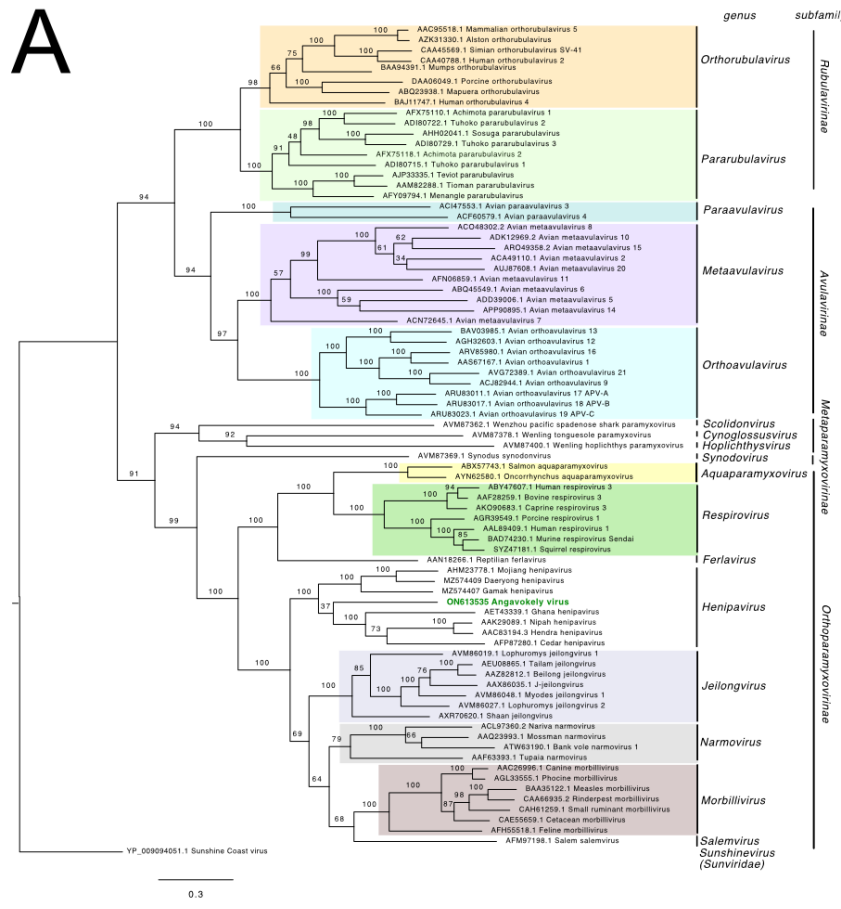
767 MZ574407; Daeryong virus (DARV) MZ574409. C. AlphaFold detail from full structure prediction
768 for AngV. Localization of ephrin binding sites conserved across most HNVs is colored green and
769 labeled corresponding to position in the NiV and AngV genomes.



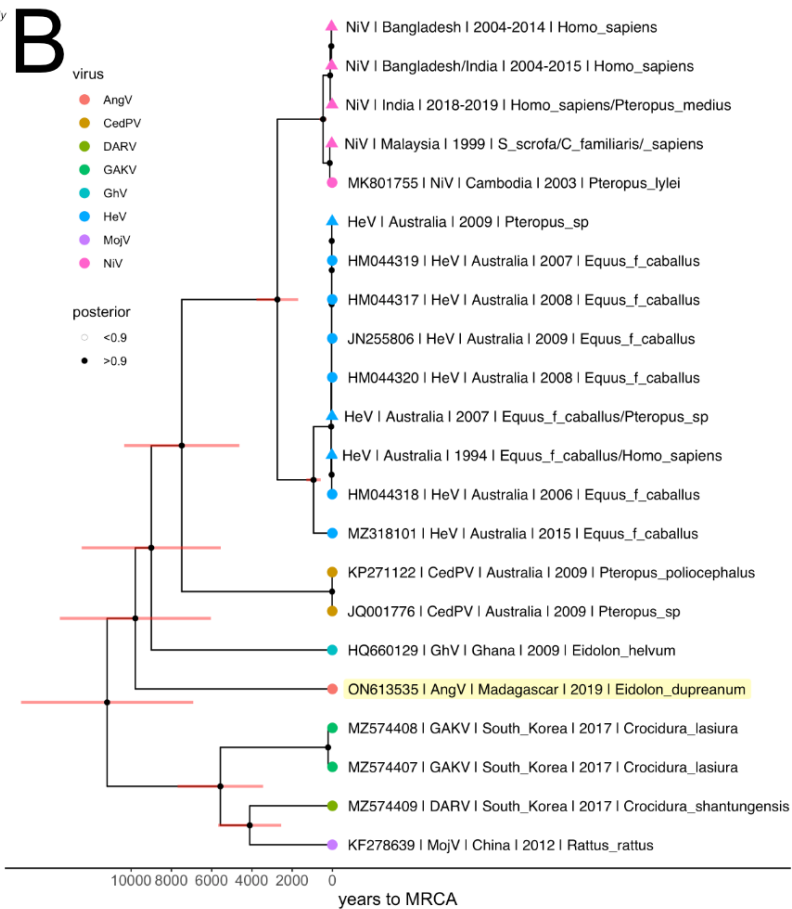


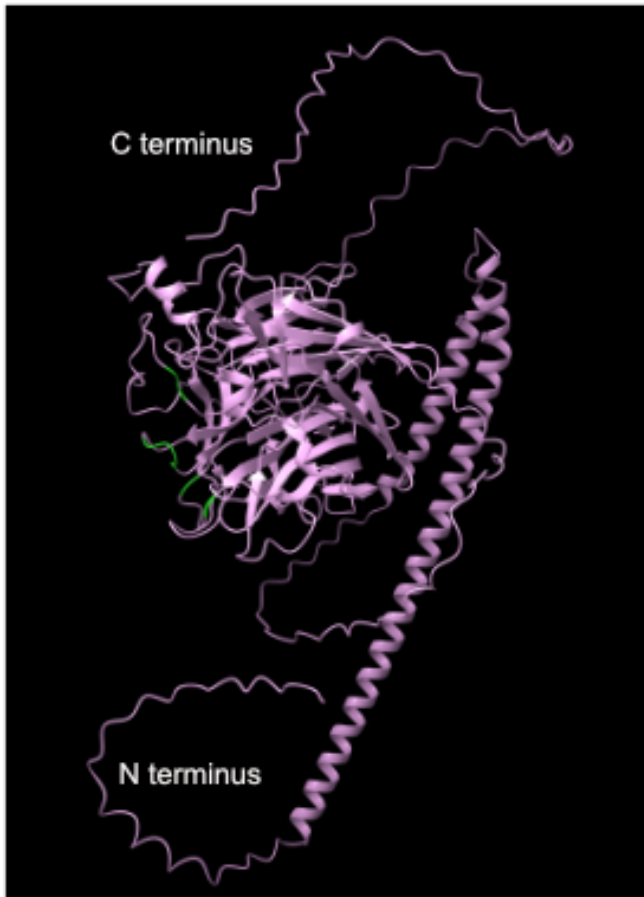


A



B



A**B**

	AngV 521 Niv 502					AngV 548 Niv 529					AngV 575 Niv 556				AngV 599 Niv 580						
			*	*			*	*	*	*		*				*					
AngV	F	C	H	S	S	G	Y	...	S	H	S	Y	G	...	I	A	A	G	...	I	R
Niv	I	C	W	E	G	V	Y	...	N	Q	T	A	E	...	T	N	A	Q	...	I	Y
HeV	V	C	W	E	G	T	Y	...	N	Q	T	A	E	...	T	N	A	Q	...	I	Y
MojV	I	C	N	T	R	G	Y	...	N	N	G	G	T	...	Y	S	I	T	...	G	K
CedV	I	C	Y	G	G	T	Y	...	D	Q	L	A	E	...	L	N	T	R	...	T	N
GhV	V	C	W	E	G	T	Y	...	E	Q	V	A	E	...	S	S	A	R	...	I	T
GAKV	S	C	K	T	W	N	F	...	K	T	G	N	S	...	Q	S	I	G	...	G	V
DARV	V	C	S	S	Y	G	Y	...	N	G	E	G	T	...	F	K	I	I	...	G	Q

C