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

Running Title Novel Henipavirus, Angavokely virus, discovered in Madagascar fruit bats

Authors

Sharline Madera[1], Amy Kistler[2], Hafaliana C. Ranaivoson[3,4,5], Vida Ahyong[2], Angelo Andrianiaina[5], Santino Andry[6], Vololoniaina Raharinosy[4], Tsiry H. Randriambolamanantsoa[4], Ny Anjara Fifi Ravelomanantsoa[5], Cristina M. Tato[2], Joseph L. DeRisi[2,7], Hector C. Aguilar[8], Vincent Lacoste[4], Philippe Dussart[4], Jean-Michel Heraud[4,9], Cara E. Brook[3]

Affiliations

[1] Division of Infectious Diseases, Department of Medicine, University of California, San Francisco, CA, USA

[2] Chan Zuckerberg Biohub, San Francisco, CA, USA

[3] Department of Ecology and Evolution, University of Chicago, Chicago, IL, USA

[4] Virology Unit, Institut Pasteur de Madagascar, Antananarivo, Madagascar

[5] Department of Zoology and Animal Biodiversity, University of Antananarivo, Antananarivo, Madagascar

[6] Department of Entomology, University of Antananarivo, Antananarivo, Madagascar

[7] Department of Medicine and Biochemistry & Biophysics, University of California, San Francisco, CA, USA

[8] Department of Microbiology and Immunology, College of Veterinary Medicine, Cornell University, Ithaca, NY, USA

[9] Virology Department, Institut Pasteur de Dakar, Dakar, Senegal

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

Abstract

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

Importance

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

88 **Introduction**

89 Henipaviruses (HNVs) belong to a genus of bat-, rodent-, and shrew-borne viruses within the
90 family *Paramyxoviridae* with demonstrated zoonotic potential. HNVs can manifest extreme
91 virulence in human hosts, as exemplified by the prototypical HNVs, Hendra virus (HeV), and
92 Nipah virus (NiV), which cause severe acute respiratory distress and/or encephalitis in humans,
93 yielding case fatality rates that can exceed 90% (1–3). This high pathogenicity and the lack of
94 approved HNV therapeutics or vaccines for humans have garnered HeV and NiV classification as
95 Biological Safety Level 4 (BSL4) agents and WHO priority diseases. Since their discovery in the
96 1990s, HeV and NiV have periodically spilled over to humans from their reservoir hosts,
97 pteropodid bats. HeV zoonosis is mediated by spillover to intermediate horse hosts, from which
98 humans acquire infection (4). NiV can spillover to humans via intermediate transmission
99 through pig hosts, or directly from bat-to-human, resulting in near-annual outbreaks of fatal
100 encephalitis in South Asia, where subsequent human-to-human transmission also occurs (2, 5–
101 7).
102
103 Novel HNVs continue to emerge from wildlife hosts and represent ongoing threats to human
104 health. Initially, the *Henipavirus* genus comprised only HeV and NiV; however, the past two
105 decades have witnessed the discovery of five new HNVs: bat-borne Cedar virus (CedV) and
106 Ghanaian bat virus (GhV), rodent-borne Mojiang virus (MojV), and shrew-borne Gamak (GAKV)
107 and Daeryong viruses (DARV) (8–11). Of these novel HNVs, at least two show evidence of
108 zoonotic potential: serological data suggests prior human exposure to GhV in West Africa (12),
109 while MojV was first identified following a human outbreak of severe pneumonia in Chinese
110 mine workers, all of whom died after infection (9). In addition to their high potential for
111 pathogenicity, HNVs possess a broad host range that spans at least seven mammalian orders,
112 including bats (10, 13).
113
114 Cross-species viral spillover necessitates effective inter-species transmission, which first
115 requires a virus to successfully enter the cells of diverse host species. In general, HNVs use the
116 highly-conserved ephrin family of proteins, both type A and type B, as cell entry receptors (1, 8,

117 14–16). A notable exception to this pattern is MojV, which does not use ephrin proteins—or the
118 sialic acid and CD150 receptors common to non-HNV paramyxoviruses—to gain cell entry (14,
119 17). Indeed, as of yet, the viral entry receptor for MojV—and the closely related GAKV and
120 DARV—remain unknown. In general, viruses in the genus *Henipavirus* have broad host ranges
121 and cause high case fatality rates following human spillover, making the characterization of new
122 HNVs a high public health priority.
123
124 The HNV genome consists of six structural proteins: nucleocapsid (N), phosphoprotein (P),
125 matrix (M), fusion (F), glycoprotein (G), and polymerase (L). In comparison with other members
126 of the family *Paramyxoviridae,* HNVs have relatively larger genomes (approximately 18kb vs
127 16kb). This extended length is largely due to several, long 3' untranslated regions (UTR) of the
128 N, P, F and G mRNAs (18, 19). The genome length of HNVs, like all paramyxoviruses, adheres to
129 the so-called 'Rule of Six', whereby viral genomes consistently demonstrate polyhexameric
130 length (20). The 'Rule of Six' is believed to be a requirement for efficient genome replication
131 under the unique mRNA editing features of the paramyxovirus genome (20). The paramyxovirus
132 P locus exhibits notable transcription properties that are shared across most members of the
133 *Paramyxoviridae* family. The P gene permits the translation of additional accessory proteins
134 from either gene editing events within the locus (prior to translation) or an overlapping ORF in
135 the P gene itself. All HNVs, with the exception of CedV, harbor a highly conserved mRNA editing
136 site at which the insertion of additional guanine residues can result in the translation of
137 accessory proteins, V and W, involved in viral antagonism and evasion of the host immune
138 system (1). The HNV P gene also contains an overlapping ORF that allows for the synthesis of a
139 third accessory protein, C, which is also involved in viral host immune evasion (1).
140
141 Our lab has previously demonstrated evidence of exposure to henipa-like viruses in serum
142 collected from three endemic Madagascar fruit bat species (*E. dupreanum, Pteropus rufus,* and
143 *Rousettus madagascariensis*) using a Luminex serological assay which identified cross reactivity
144 to CedV/NiV/HeV-G and -F proteins (21). The most significant antibody binding previously
145 detected corresponded to the NiV-G antigen for *E. dupreanum* serum and the HeV-F antigen for

146 *P. rufus* and *R. madagascariensis* serum, suggesting the potential circulation of multiple HNVs in
147 the Madagascar fruit bat system (21). Fruit bats, including *E. dupreanum,* are consumed widely
148 in Madagascar as a source of human food, presenting opportunities for cross-species zoonotic
149 emergence. This underscores the importance of further characterization of the pathogenic and
150 zoonotic potential of AngV and other potential HNVs circulating in the Madagascar fruit bat
151 system. Here, we describe and characterize a novel bat HNV, Angavokely virus (AngV),
152 recovered from urine samples collected from the Madagascar fruit bat, *E. dupreanum*. Our
153 work suggests AngV is part of an ancestral group of HNVs that may rely on a novel, non-ephrin-
154 mediated viral entry pathway.
155
156 **Methods**
157 *Ethics Statement*
158 Animal capture and handling and subsequent collection of biological samples were conducted
159 in strict accordance with the Madagascar Ministry of Forest and the Environment (permit
160 numbers 019/18, 170/18, 007/19) and guidelines posted by the American Veterinary Medical
161 Association. Field protocols were approved by the UC Berkeley Animal Care and Use Committee
162 (ACUC Protocol # AUP-2017-10-10393), as previously described (22).
163
164 *Animal capture, sample collection, and RNA extraction*
165 Fruit bats were captured and processed in part with a long-term study investigating the
166 seasonal dynamics of potentially zoonotic viruses in Madagascar, as has been previously
167 described (21–25). Animals were identified morphologically by species, sex, and age class
168 (juvenile vs. adult), and urine swabs were collected into viral transport medium from any
169 individual that urinated during handling. Urine swabs were flash-frozen in liquid nitrogen in the
170 field and delivered to -80^{0}C freezers at Institut Pasteur de Madagascar for long-term storage.
171 Urine specimens from 206 bats were randomly selected for total RNA extraction using the Zymo
172 Quick RNA/DNA Microprep Plus kit, performed as previously described (22).
173
174 *mNGS library preparation*

175 Total urine RNAs were diluted with nuclease-free H_2O, and 5uL of each specimen was used as
176 input for mNGS library preparation. A 2-fold dilution series of a 25ng/uL stock of HeLa total RNA
177 (n=8 samples), along with 5 water samples were included and processed in parallel as positive
178 and negative controls, respectively. Additionally, a 25pg aliquot of External RNA Control
179 Consortium (ERCC) spike-in mix (Thermo-Fisher) was included in each sample. Dual-indexed
180 mNGS library preparations for the samples were miniaturized and performed in 384-well
181 format with NEBNext Ultra II RNAseq library preparation kit (New England Biolabs) reagents.
182 RNA samples were fragmented for 12 min at 94°C, and 16 cycles of PCR amplification were
183 performed. Per sample read yields from a small scale iSeq (Illumina) paired-end 2 x 146bp
184 sequencing run on an equivolume pool of the individual libraries were used to normalize
185 volumes of the individual mNGS libraries to generate an equimolar pool. Paired-end 2 x 146bp
186 sequencing of the resulting equimolar library pool was performed on the NovaSeq6000
187 (Illumina) to obtain approximately 50 million reads per sample.
188
189 S*equence analysis*
190 Raw reads from urine sample sequencing were first uploaded to the CZBID (v6.8) platform for
191 host and quality filtering and *de novo* assembly (26). In brief, in the CZBID pipeline, adaptor
192 sequences were removed with Trimmomatic (v.0.38), and reads were quality filtered (27).
193 Reads then underwent host-filtration against the Malagasy fruit bat genome, *E. dupreanum,*
194 using STAR (v 2.7.9a) (28) and a second host-removal step using Bowtie2 (29). After host
195 filtering, reads were aligned using rapsearch2 (30) and Gsnap (31), and putative pathogen taxa
196 were identified. Next, reads were assembled using SPADES (v.3.15.3) (32), and all contigs
197 generated were subject to BLAST analysis against the putative taxa previously identified by
198 rapsearch2 and Gsnap. We considered samples positive for HNV if the CZBID pipeline produced
199 at least one contig with an average read depth of two or more, which yielded a BLAST
200 alignment length >100 nt/aa and an e-value < 0.00001 (BLASTn v2.5.0+) or a bit score >100
201 (BLASTx v2.5.0+) when queried against an HNV database derived from all HNV genomes
202 available in NCBI (Accessed July 2021).
203

204 *Genome Annotation and Comparison*

205 One urine sample, collected from an adult female *E. dupreanum* fruit bat in March 2019,
206 yielded a near full-length HNV genome, which we analyzed in greater depth in subsequent
207 analyses and annotated as the novel HNV, AngV (Genbank Accession #: ON613535). Nucleotide
208 BLAST of the AngV genome identified NiV (GenBank Accession #: AF212302) as the top hit for
209 this novel virus and was subsequently chosen as the reference genome for further analysis. We
210 aligned AngV to NiV (GenBank Accession #: AF212302) in the program Geneious Prime
211 (v2020.2.4) and annotated all six major HNV structural genes, and the accessory C ORF, within
212 the P gene. We identified the putative mRNA editing site within the P gene sequence (spanning
213 nucleotides 1,225-1,232 of the P gene) and manually added one or two guanine (G) residues to
214 the 3' end of the conserved HNV mRNA editing site to generate V and W ORFs, respectively,
215 and their corresponding proteins. We furthered queried all identified transcriptional elements
216 against publicly available sequences using NCBI BLAST and BLASTx (33). Resulting BLAST and
217 BLASTx hits were used in phylogenetic analyses as described below.

218

219 We used the program pySimPlot to scan the whole genome sequence of AngV for nucleotide
220 sequence identity to the NiV genome (GenBank Accession #: AF212302) and the nucleotide and
221 amino acid sequences of individual Open Reading Frames (ORFs) contained therein.
222 Respectively, window size and scanning were specified as 50 and one for nucleotide pairwise
223 identity and 50 and five for amino acid pairwise identity. Results were visualized using Prism
224 (9.2.0).

225

226 *Phylogenetic analyses*

227 We constructed 10 Maximum Likelihood (ML) phylogenetic trees to analyze the evolutionary
228 relatedness of our putative HNV to previously described paramyxoviruses. These included (a)
229 one amino acid L protein phylogeny comparing the L protein of AngV to all reference L protein
230 paramyxovirus sequences in NCBI, in addition to the newly-described shrew HNV, GAKV and
231 DARV (accessed November 2021) and (b) nine amino acid phylogenies comparing each
232 individual protein annotated in AngV (N, P, C, V, W, M, F, G, L) against the top 50 BLASTx

233 sequence hits for each protein collapsed on 98% sequence similarity. Distinct outgroups were
234 applied: (a) Sunshine Coast Virus (GenBank Accession #: YP_009094051.1) for L protein, (b)
235 Human orthopneumovirus (HRSV, GenBank Accession #: NC_001781) for N, P, M, F, G, L
236 proteins, and (c) Sendai virus (GenBank Accession #: NP_056872) for gene C.

237 For each phylogenetic tree, we aligned sequences via the MUSCLE algorithm (v3.8.1551) (34)
238 and determined the best fit nucleotide or amino acid substitution model for using ModelTest-
239 NG (35). Phylogenies were then constructed in RAxML-NG (36), using the corresponding best fit
240 model: JTT (complete L-protein sequence) or LG+G4+F (individual proteins). In accordance with
241 best practices outlined in the RAxML-NG manual, twenty ML inferences were made on each
242 original alignment. Bootstrap replicate trees were inferred using Felsenstein's method (37).
243 MRE-based bootstopping test was applied after every 200 replicates (38), and bootstrapping
244 was terminated once diagnostic statistics dropped below the threshold value. Bootstrap
245 support values were drawn on the best-scoring tree.

246 We additionally computed one Bayesian time-resolved phylogeny, using all 77 full-length HNV
247 nucleotide sequences available in NCBI, including our newly contributed AngV (GenBank
248 Accession #: ON613535). As with ML trees, sequences were first aligned in MUSCLE (v3.8.1551)
249 (34), and the best fit nucleotide substitution model was subsequently queried in ModelTest-NG
250 (35). We then constructed a Bayesian timetree in the program BEAST2 (39, 40), using the best
251 fit GTR+I+G4 model inferred for the whole genome alignment from ModelTest-NG and
252 assuming a constant population prior. Sampling dates corresponded to collection data as
253 reported in NCBI Virus; we assumed a collection date of 31-July in cases where only year of
254 collection was reported. We computed trees using both an uncorrelated exponentially
255 distributed relaxed molecular clock (UCED) and a strict clock but here report results from the
256 strict clock only as similar results were inferred from both. We ran Markov Chain Monte Carlo
257 (MCMC) sample chains for 1 billion iterations, checked convergence using TRACER v1.7 (41) and
258 averaged trees after 10% burn-in using TreeAnnotator v2.6.3 (42) to visualize mean posterior
259 densities at each node. The resulting phylogeny was visualized in R v.4.0.3 for MacIntosh in the
260 'ggtree' package (43).

261

262 *AngV G Protein Structure Modeling*

263 We used the Artificial Intelligence system, AlphaFold, to predict the 3D structure of the AngV
264 glycoprotein (G) (44). Molecular graphics and analyses of the AngV glycoprotein structure were
265 performed with UCSF ChimeraX (45). HNV glycoprotein ephrin binding residues were aligned
266 using the program Geneious Prime (v2020.2.4).

267

268 **Results**

269 *Discovery and prevalence of HNV in Malagasy bats*

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

278

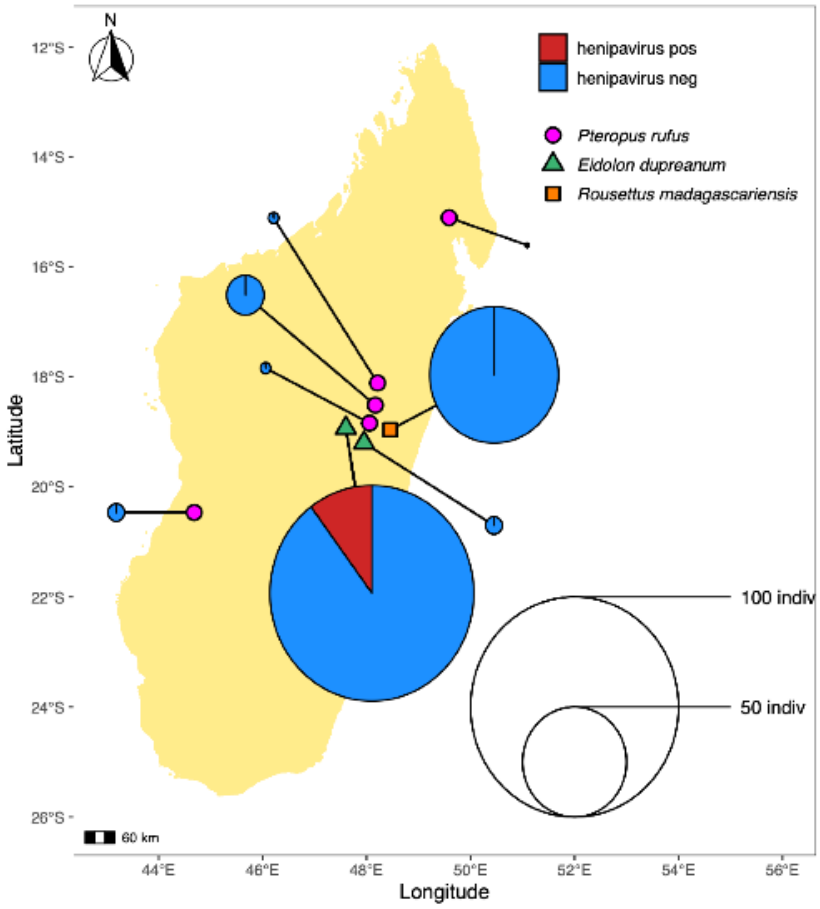
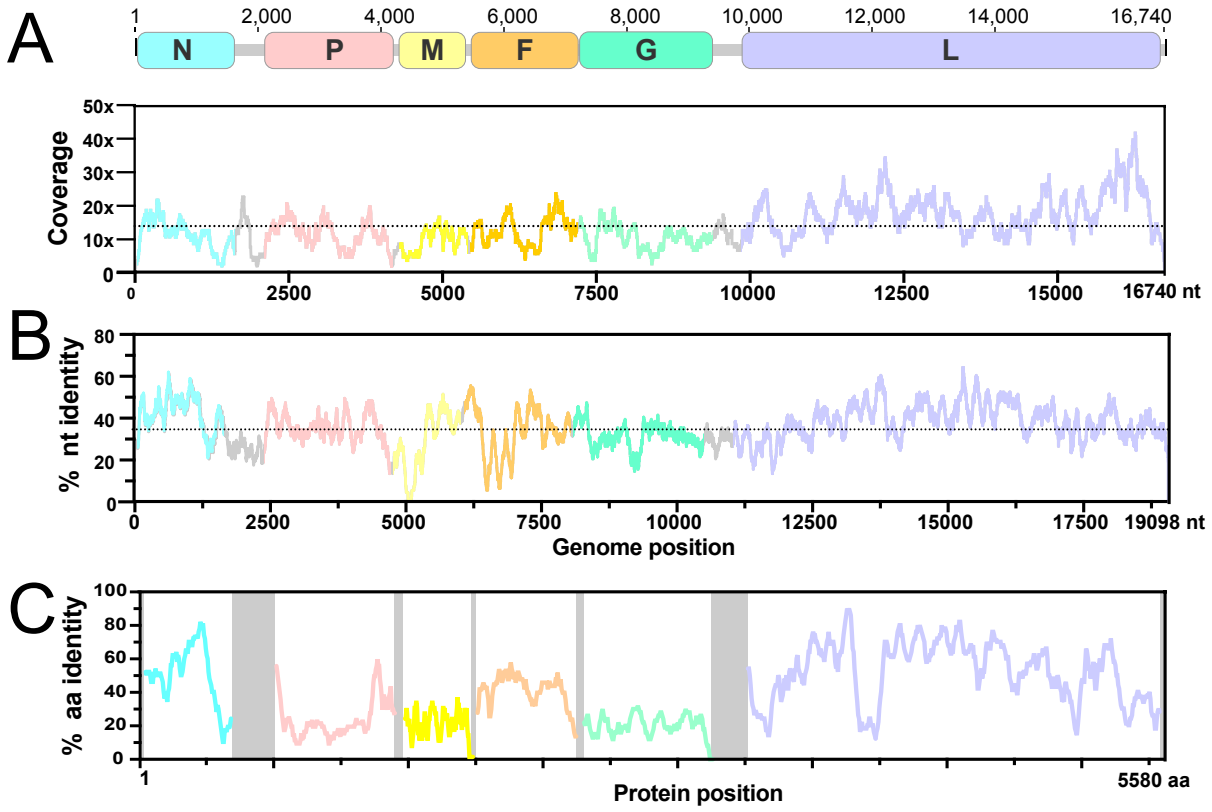


Figure 1. Geographic location of sampling sites used in this study. Sampling sites grouped by bat species found depicted as follows *P. rufus* (pink circles) Ambakoana (−18.51 S, 48.17 E) / Mahabo (-20.46 S, 44.68 E) / Mahialambo (-18.11 S, 48.21 E) / Makira (-15.11 S, 49.59 E) / Marovitsika (-18.84 S, 48.06 E) roosts; *E. dupreanum* (green triangles) Angavobe (−18.94 S, 47.95 E) /Angavokely (−18.93 S, 47.76 E) caves; *R. madagascariensis* (orange squares) Maromizaha cave (−18.96 S, 48.45). Pie charts are size-weighted by total bat population sampled at each site, corresponding to the legend. The percentage of HNV positive samples is shown for all sampled species and sites. HNV positive samples were only recovered from the *E. dupreanum* Angavokely site.

Genomic characterization of AngV

We recovered one near-full-length HNV contig (16,740 nt), supported by an average sequencing depth of 14 reads (Figure 2A), from a urine sample collected from an adult, non-lactating *E. dupreanum* female in the 2018-2019 wet season (capture date: 15-March 2019). We focused

295 subsequent genomic analyses on this longest sequence, which we named Angavokely virus
296 (AngV) after the site of *E. dupreanum* capture.
297



298

299 **Figure 2.** AngV genome organization. A. Coding regions for each gene are shown and depicted
300 in color, non-coding intergenic and terminal regions are highlighted in gray. Depicted genes
301 represented as follows: nucleocapsid (N), RNA polymerase (P and L), matrix (M), fusion (F), and
302 glycoprotein (G). Sequencing read depth supporting each position of the recovered genome
303 sequence is plotted below the genomic schematic. Scanning nucleotide (B) and amino acid (C)
304 pairwise identity to Nipah virus (GenBank Accession #: AF212302). Dotted horizontal lines
305 represent average read depth (14.29) or average nucleotide pairwise identity (36%).
306

307

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

AngV coding regions

The P gene of AngV follows an organization similar to most members of the *Henipavirus* genus. AngV harbors alternative start sites which, respectively, encode the P and C proteins, as well as a conserved putative mRNA editing site common to most paramyxoviruses A[4-6]G[2-3] (Figure 3A and B). AngV shares an identical putative mRNA editing site with the recently discovered HNVs, MojV and GAKV. Pseudotemplated addition of one or two G residues at the conserved putative mRNA editing site generates a putative V and W protein, respectively (Figure 3C; W protein in Supplemental Figure 1). In congruence with members of the *Henipavirus* genus that encode the conserved putative mRNA editing site, the putative V protein of AngV harbors a unique C-terminal region that contains a highly conserved cysteine-rich zinc finger domain (Figure 3C).

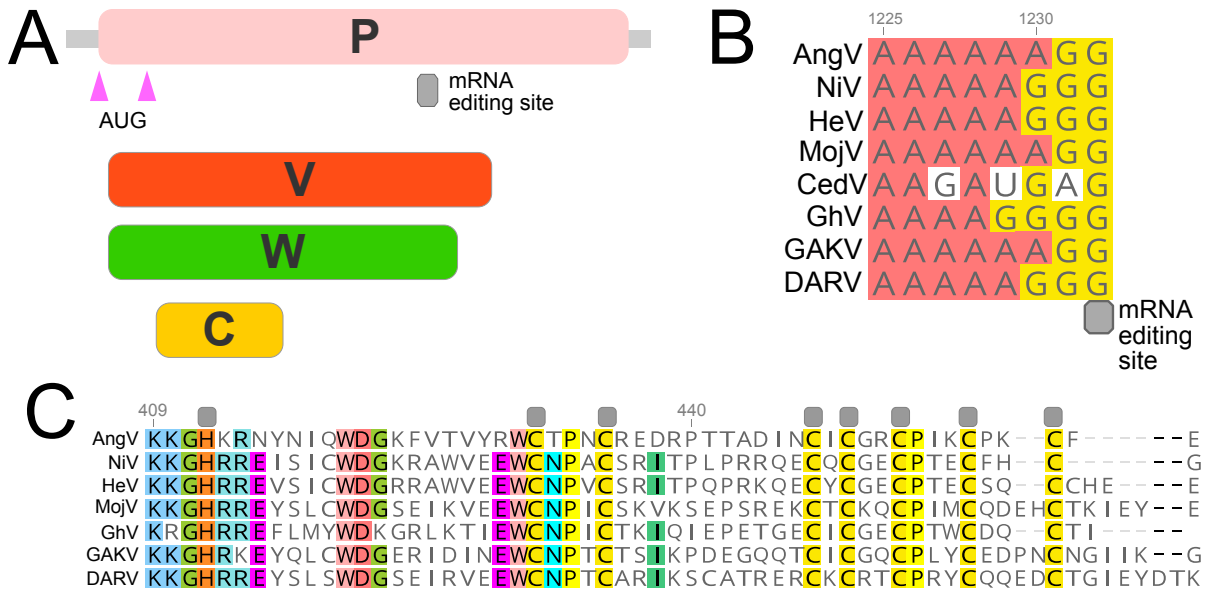


Figure 3. Organization of the P gene of AngV. A. Alternative transcriptional start sites (pink triangle) generate the P and C protein. Pseudotemplated addition of one or two guanine nucleotides at the putative mRNA editing site generates a V and W protein, respectively. B. Sequence alignment of the putative mRNA editing site across members of the *Henipavirus* genus (cRNA depicted). C. Amino acid alignment of the unique C terminal region of the V protein following the addition of one guanine nucleotide to the putative mRNA editing site. Gray boxes denote conserved cysteine and histidine residues suggested to directly coordinate bound zinc ions (54). Individual nucleotides or amino acids are color coordinated if at least 75%

335 conserved at the alignment position. Nucleotide or amino acid position numbers displayed
336 represent the position within the AngV gene or protein. Virus name (abbreviation), followed by
337 GenBank Accession #: Angavokely virus (AngV) ON613535; Nipah virus (NiV) AF212302; Hendra
338 virus (HeV) AF017149; Mojiang virus (MojV) KF278639; Cedar virus (CedV) JQ001776; Ghanaian
339 bat Henipavirus (GhV) HQ660129; Daeryong virus (DARV) MZ574409; Gamak virus (GAKV)
340 MZ574407. CedV was not included in panel C alignment because it does not encode a V protein.
341
342

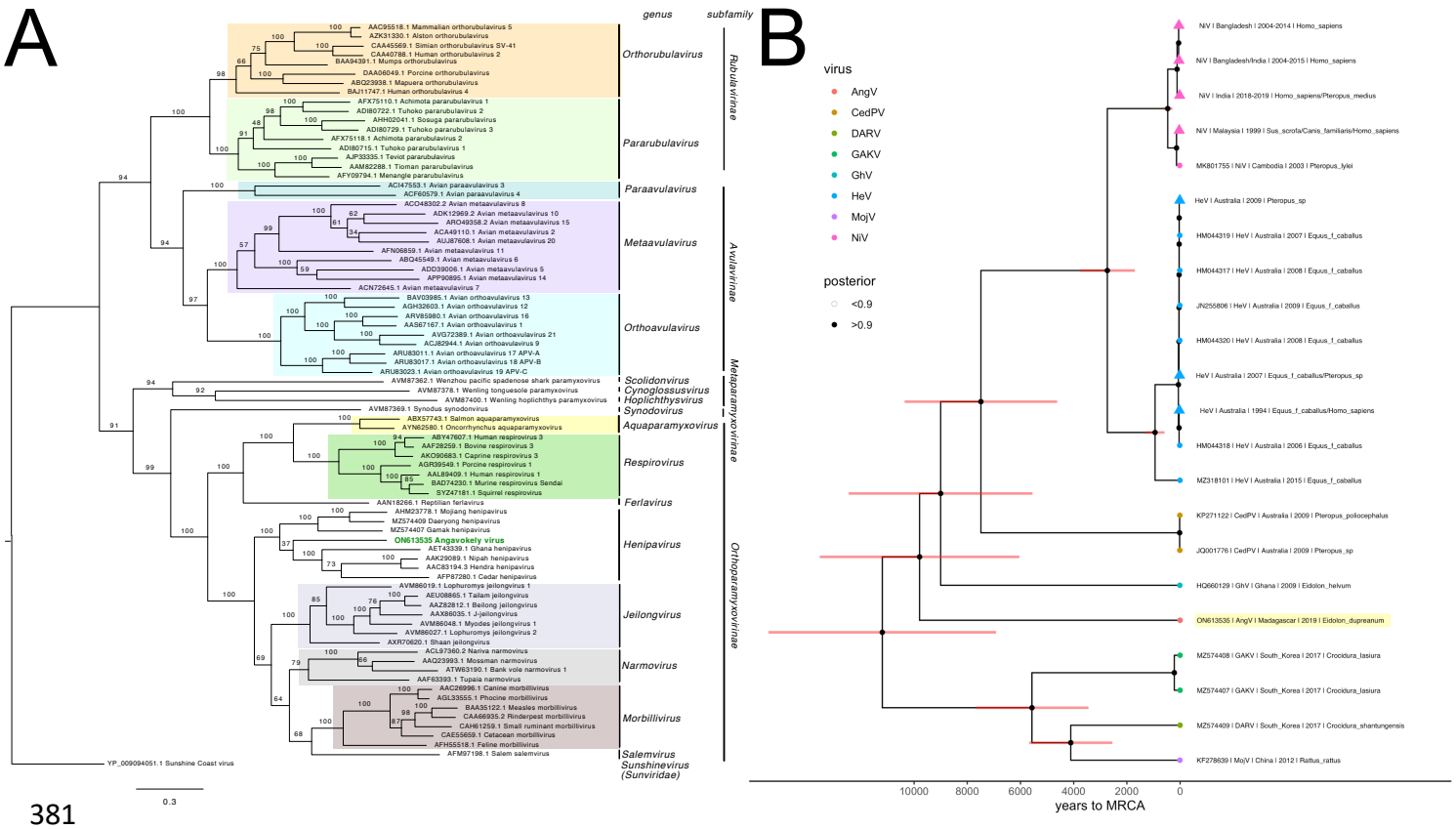
343 The length of each ORF in the AngV genome resembles those from previously described HNVs,
344 with the exception of the gene encoding the glycoprotein (G)—which, at 688 aa, is 56 aa longer
345 than the longest previously-characterized HNV glycoprotein from GhV (632 aa; Table 2) (15).
346 BLAST analysis indicates that AngV ORFs for genes encoding the nucleocapsid (N), matrix (M),
347 and polymerase (L) proteins exhibit the highest nucleotide and amino acid pairwise identity
348 with other HNVs, with highest similarity shared with the NiV L protein (nt 74.7%, aa 52.2%,
349 Table 2). In contrast to many emerging viruses, AngV largely exhibits higher nucleotide vs.
350 amino acid identity with other HNVs (Table 2). The more recently discovered HNVs (MojV,
351 CedV, GAKV, DARV, GhV) mirror this pattern, showing higher nucleotide vs. amino acid identity
352 when compared to NiV and HeV (data not shown).
353

354 *AngV non-coding regions*

355 Examination of all viral intergenic regions (in cRNA orientation) reveals that AngV exhibits the
356 highly conserved CTT intergenic junction site characteristic of other HNVs, as well as gene stop
357 and gene start sites with high similarity to those of previously described HNVs (Supplemental
358 Table 1). We were unable to locate the intergenic junction site and transcriptional start or stop
359 site in the 5' region of the N ORF for AngV, suggesting that the genomic 3' untranslated regions
360 (UTR) for AngV have not yet been fully recovered. Comparison of the 5' and 3' UTRs for AngV
361 with those of other HNVs reveals UTRs of varying lengths within the *Henipavirus* genus
362 (Supplemental Table 2). Nevertheless, AngV exhibits similar lengths and a nucleotide identity of
363 roughly 30-40% for the 5' and 3' UTRs of most HNV genes; however, the P gene 3' UTR, the M
364 gene 5' and 3' UTRs, and the F gene 5' and 3' UTRs, are significantly shorter in AngV compared
365 with previously described HNVs. Correspondingly, nucleotide identity varies when comparing
366 this shorter subset of 5' and 3' UTRs for AngV against other HNVs (Supplemental Table 2).

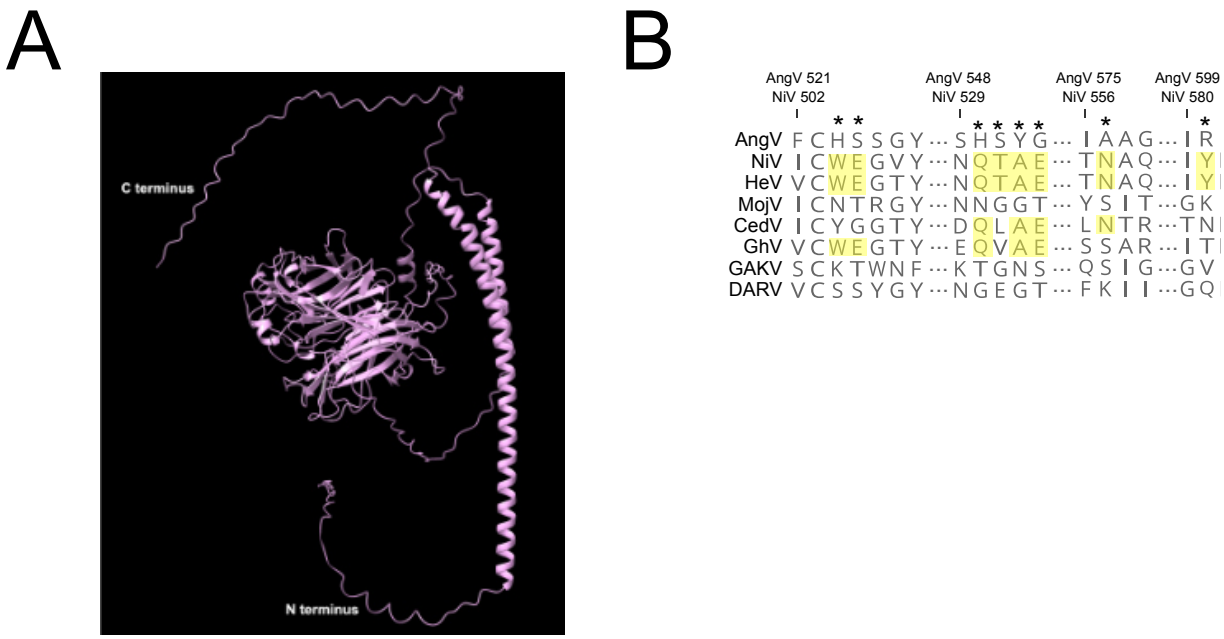
Phylogenetic Analyses

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



382 **Figure 4.** A. Phylogenetic analysis of the complete L protein sequences of members of the
383 family *Paramyxoviridae*. Tree is rooted with Sunshine Coast Virus (GenBank Accession #:
384 YP_009094051.1) as an outgroup, with outgroup branch length shrunk for ease of viewing.
385 Novel HNV, AngV, is depicted in green. Subfamilies and genera are demarcated, excluding those
386 unassigned to subfamily (genera *Scolidonvirus, Cynoglossusvirus, Hoplichthysvirus*). Bootstrap
387 support is depicted and GenBank Accession numbers displayed next to virus names. Scale bar
388 represents substitutions per site. B. Time-resolved Bayesian phylogeny computed in BEAST2
389 incorporating all available *Henipavirus* whole genome nucleotide sequences, with the addition
390 of newly discovered GAKV, DARV, and AngV. Closely-related sequences are collapsed at triangle
391 nodes for NiV and HeV (phylogeny with un-collapsed branches available in Supplemental Figure
392 2). 95% HPD intervals around the timing of each branching node are visualized as red horizontal
393 bars. Posterior support >.9 is indicated by black coloring of the corresponding node, and distinct
394 *Henipavirus* species are indicated by colored tip points, with AngV highlighted in yellow for
395 further emphasis. The estimated time to MRCA for Angavokely virus and the previously-
396 described bat-borne HNVs is 9,794 (95% HPD 6,519 – 14,025) years ago.
397
398
399 In N, P, C, V, W, M and G amino acid phylogenies, the AngV proteins cluster closely with those
400 of other HNVs (Supplemental Figure 3). Interestingly, in the amino acid phylogeny, the AngV F
401 protein, like the F proteins of MojV, GAKV, and DARV, localizes ancestral to non-HNV
402 paramyxoviruses and distinct from the bat-borne HNV clade (Supplemental Figure 3). The AngV
403 L protein shows the highest amino acid identity to the L protein of rodent-borne Mount Mabu
404 Lophuromys virus 2 (MMLV-2), a putative Jeilongvirus (46), but is nonetheless nested between
405 the MojV/GAKV/DARV clade and the bat-borne HNV clade (Supplementary Figure 3).
406
407 *AngV glycoprotein*
408 We further examined the AngV G protein for conserved structural features and amino acid
409 residues historically associated with HNV ephrin binding. AlphaFold analysis revealed a six-
410 bladed β-propeller fold that is characteristic of *Paramyxoviridae* glycoproteins, with each blade
411 largely composed of 4 antiparallel β-strands (Figure 5A). The β-propeller fold is stabilized by
412 seven disulfide bonds that are conserved among HNVs (Supplemental Figure 4). This HNV
413 protein G structure-based alignment reveals that the elongated AngV G protein primarily
414 results from a lengthy C-terminal tail with an additional 67 aa beyond that of NiV G protein

415 (Supplemental Figure 4). Similar to the MojV G protein, the AngV G protein lacks previously
416 described ephrin binding residues (NiV W504, E505, T531, A532, E533, N557, and Y581) (Figure
417 5B and Supplemental Figure 4) (16, 47, 48).
418



419

420 **Figure 5.** AlphaFold-predicted AngV glycoprotein 3D structure and ephrin binding residue
421 sequence alignment. A. AlphaFold-predicted 3D structure of AngV glycoprotein. N and C
422 termini are indicated in white. B. Alignment of HNV ephrin binding residues. The position of
423 previously-described HNV ephrin binding residues are noted by a star, and residues conserved
424 across most HNVs are highlighted yellow. Amino acid position numbers displayed represent the
425 position within the AngV or NiV glycoproteins. Virus name (abbreviations) followed by GenBank
426 Accession #: Angavokely virus (AngV) ON613535; Nipah virus (NiV) AF212302; Hendra virus
427 (HeV) AF017149; Mojiang virus (MojV) KF278639; Cedar virus (CedV) JQ001776; Ghanaian bat
428 Henipavirus (GhV) HQ660129; Gamak virus (GAKV) MZ574407; Daeryong virus (DARV)
429 MZ574409.
430

431

432 **Discussion**

433 We describe and characterize a novel HNV, AngV, from a urine sample collected from an *E.*
434 *dupreanum* Malagasy fruit bat. In this study, urine samples from 206 unique fruit bats were
435 assessed by metagenomic sequencing, yielding an overall positive HNV detection rate of 4.9%

436 (10/206) for all bats studied and a HNV prevalence of 9.4% (10/106) for the *E. dupreanum*
437 hosts. Of all the HNV positive samples, only one sample yielded sufficient reads for assembly of
438 a complete coding sequence and subsequent genomic analysis. In a 6-year collection period
439 spanning multiple wet/dry seasons, HNV positive samples were only recovered from *E.*
440 *dupreanum* bats in the Angavokely roosting site, despite prior serological evidence of HNV
441 infection in *P. rufus* and *R. madagascariensis* bats, as well (21). HNV RNA was recovered from *E.*
442 *dupreanum* in both wet and dry seasons, though higher sampling intensity throughout the wet
443 season precludes any conclusions regarding underlying seasonal patterns in these data.
444 Previous work in this system has suggested a seasonal increase in fruit bat seroprevalence
445 across the winter low nutrient season, which also overlaps the gestation period for these
446 synchronously breeding fruit bats (21). In fruit bat systems elsewhere, HNVs are also shed in
447 urine at higher rates during the nutrient-poor dry seasons for the localities in question (49–51);
448 in the case of NiV and HeV, these seasonal viral shedding pulses have been linked to zoonotic
449 spillover.
450
451 The recovered genome of AngV exhibits a structural organization characteristic of the
452 *Henipavirus* genus and a nucleotide and amino acid identity to HeV and NiV that is comparable
453 to those shared with the more distantly related HNVs, MojV, GhV and CedV. A limited quantity
454 of available original sample precluded full genome recovery for AngV (as evidenced by the lack
455 of the 5' UTR region of the N ORF), which prevented analysis of the extent to which the full
456 AngV genome may abide by the 'Rule-of-Six', observed by all other members of the
457 *Orthoparamyxovirinae* subfamily (52). Phylogenetic analyses of AngV support classification of
458 this virus as a distinct novel bat-borne *Henipavirus* (L gene amino acid distance <0.82 distance
459 for the subfamily *Orthoparamyxovirinae*), in accordance to the International Committee on
460 Taxonomy of Viruses (ICTV) criteria (19). This novel HNV is estimated to have diverged
461 approximately 9,800 years ago, prior to the currently known African and Asian bat-borne HNV
462 lineages but considerably more recently than the estimated mid-to late-Miocene divergence of
463 *E. dupreanum* from its sister species, *E. helvum,* on the African continent (53). Recent
464 characterization of *Betacoronaviruses* in Madagascar fruit bats demonstrates surprising identity

465 to lineages circulating in West Africa (22), suggesting that, despite their endemism, Malagasy
466 fruit bats likely experience some form of contact with the African continent. Of the 49 bat
467 species that inhabit the island nation of Madagascar, nine species are widely distributed across
468 Africa, Asia, and/or Europe, presenting opportunities for inter-species viral transmission via
469 island-hopping. Intensified viral sampling of Madagascar's insectivorous bat populations for
470 HNVs thus represents an important future research priority.
471
472 As an ancestral bat-borne HNV, AngV may provide important insight into HNV evolution and
473 pathogenesis. Similar to other paramyxoviruses, the encoded AngV P gene is able to produce
474 multiple immunomodulatory protein products (54). One such protein product is the V protein,
475 thought to be involved in immune evasion and considered a significant determinant of viral
476 pathogenicity and lethality (55, 56). AngV harbors the highly conserved mRNA editing site and a
477 predicted ORF that encodes a V protein with a conserved cysteine-rich C-terminus, suggesting
478 that AngV has the capacity to produce a functional V protein. With the exception of the newly
479 discovered HNVs in shrews, GAKV and DARV, all HNVs harboring a V protein have previously
480 demonstrated evidence of human infection, highlighting the potential for AngV to cause
481 productive infection in humans (1, 9, 12). Further studies are needed to ascertain the virulence
482 potential and host breadth of this novel virus.
483
484 Characterization of the AngV glycoprotein (G) through AlphaFold modeling and structure-based
485 alignments revealed a similar structural organization to other HNV glycoproteins. Notably, the
486 AngV glycoprotein surpasses that of GhV as the longest glycoprotein of the *Henipavirus* genus.
487 Like that of GhV, the AngV glycoprotein harbors a long C terminal extension (Supplemental
488 Figure 4). It is unclear if the C terminal extension of the AngV glycoprotein has a functional role,
489 though the C terminal extension of the glycoprotein in GhV is known to play a functional role in
490 receptor-mediated fusion (15).
491
492 Henipavirus host tropism and virulence rely on a myriad of factors, one of which is the HNV
493 glycoprotein. The previously characterized HNV glycoproteins of NiV, HeV, CedV, and GhV,

494 utilize members of the ephrinA and ephrinB class family as host-cell receptors for viral entry
495 into human cells (15–17, 47, 57). However, like MojV, the AngV glycoprotein lacks these well-
496 conserved ephrin binding residues. Structure-based alignments can shed light on potential
497 receptor binding residues when characterizing novel viruses. For instance, sequence-based
498 comparisons of the GhV and NiV glycoproteins were used to predict GhV ephrin binding (12),
499 which was later confirmed by crystallography (15). Structure-based alignment of the AngV
500 glycoprotein shows a lack of highly conserved ephrin binding residues, including NiV E533 – a
501 seminal residue for ephrinB2 binding that is conserved across all ephrin binding HNVs. This
502 suggests that, like MojV—and probably DARV and GAKV—the AngV glycoprotein may not bind
503 ephrins, pointing to the possible use of an ancestral viral entry pathway. The growing number
504 of novel HNVs that appear not to rely on ephrin binding for cellular entry could warrant re-
505 evaluation of the existing HNV genus to better reflect conserved function and pathobiology.
506
507 This work presents a novel bat-HNV, AngV, identified from a Malagasy fruit bat. AngV joins a
508 growing group of ancestral HNVs with unknown cell-entry receptors. Discovery of the cell
509 surface receptor for AngV represents an important future research priority that will shed light
510 on the breadth of host range for this virus, including its zoonotic potential.
511
512
513
514
515
516
517
518
519
520
521
522

Acknowledgements

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

Funding

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

Conflict of Interests

The authors declare no competing interests.

Data Availability

Raw and assembled sequencing data are deposited in NCBI Bioproject PRJNA837298. The full genome of AngV is available in GenBank under Accession # ON613535. All raw data and code for figures can be obtained in our open-access GitHub repository: https://github.com/brooklabteam/angavokely-virus

552 **References**

553 1. Eaton BT, Broder CC, Middleton D, Wang L-F. 2006. Hendra and Nipah viruses:
554 Different and dangerous. Nat Rev Microbiol 4:23–35.
555 2. Sharma V, Kaushik S, Kumar R, Yadav JP, Kaushik S. 2019. Emerging trends of
556 Nipah virus: A review. Reviews in Medical Virology 29.
557 3. Arunkumar G, Chandni R, Mourya DT, Singh SK, Sadanandan R, Sudan P, Bhargava
558 B. 2019. Outbreak investigation of Nipah Virus Disease in Kerala, India, 2018.
559 Journal of Infectious Diseases 219:1867–1878.
560 4. Murray K, Selleck P, Hooper P, Hyatt a, Gould a, Gleeson L, Westbury H, Hiley L,
561 Selvey L, Rodwell B. 1995. A morbillivirus that caused fatal disease in horses and
562 humans. Science 268:94–7.
563 5. Hsu VP, Hossain MJ, Parashar UD, Ali MM, Ksiazek TG, Kuzmin I, Niezgoda M,
564 Rupprecht C, Bresee J, Breiman RF. 2004. Nipah virus encephalitis reemergence,
565 Bangladesh. Emerging Infectious Diseases 10:2082–2087.
566 6. Gurley ES, Montgomery JM, Hossain MJ, Bell M, Azad AK, Islam MR, Molla MAR,
567 Carroll DS, Ksiazek TG, Rota PA, Lowe L, Comer JA, Rollin P, Czub M, Grolla A,
568 Feldmann H, Woodward JL, Breiman RF. 2007. Person-to-Person transmission of
569 Nipah virus in a Bangladeshi community. Emerging Infectious Diseases 13:1031–
570 1037.
571 7. Luby SP, Gurley ES. 2012. Epidemiology of Henipavirus Disease in Humans, p. 25–
572 40. *In* Lee, B, Rota, PA (eds.), Henipavirus. Springer Berlin Heidelberg, Berlin,
573 Heidelberg.
574 8. Marsh GA, de Jong C, Barr JA, Tachedjian M, Smith C, Middleton D, Yu M, Todd S,
575 Foord AJ, Haring V, Payne J, Robinson R, Broz I, Crameri G, Field HE, Wang L-F.
576 2012. Cedar virus: a novel Henipavirus isolated from Australian bats. PLoS
577 Pathogens 8:e1002836.
578 9. Wu Z, Yang L, Yang F, Ren X, Jiang J, Dong J, Sun L, Zhu Y, Zhou H, Jin Q. 2014.
579 Novel Henipa-like virus, Mojiang paramyxovirus, in rats, China, 2012. Emerging
580 Infectious Diseases 20:1064–1066.

581 10. Lee SH, Kim K, Kim J, No JS, Park K, Budhathoki S, Lee SH, Lee J, Cho SH, Cho S, Lee
582 GY, Hwang J, Kim HC, Klein TA, Uhm CS, Kim WK, Song JW. 2021. Discovery and
583 genetic characterization of novel paramyxoviruses related to the genus
584 henipavirus in crocidura species in the republic of Korea. Viruses 13.
585 11. Drexler JF, Corman VM, Müller MA, Maganga GD, Vallo P, Binger T, Gloza-Rausch
586 F, Rasche A, Yordanov S, Seebens A, Oppong S, Adu Sarkodie Y, Pongombo C,
587 Lukashev AN, Schmidt-Chanasit J, Stöcker A, Carneiro AJB, Erbar S, Maisner A,
588 Fronhoffs F, Buettner R, Kalko EK v, Kruppa T, Franke CR, Kallies R, Yandoko ERN,
589 Herrler G, Reusken C, Hassanin A, Krüger DH, Matthee S, Ulrich RG, Leroy EM,
590 Drosten C. 2012. Bats host major mammalian paramyxoviruses. Nat Commun
591 3:796.
592 12. Pernet O, Schneider BS, Beaty SM, LeBreton M, Yun TE, Park A, Zachariah TT,
593 Bowden T a., Hitchens P, Ramirez CM, Daszak P, Mazet J, Freiberg AN, Wolfe ND,
594 Lee B. 2014. Evidence for henipavirus spillover into human populations in Africa.
595 Nature Communications 5:5342.
596 13. Amaya M, Broder CC. 2020. Vaccines to emerging viruses: Nipah and Hendra.
597 Annual Review of Virology 7:447–473.
598 14. Cheliout Da Silva S, Yan L, Dang H v., Xu K, Epstein JH, Veesler D, Broder CC. 2021.
599 Functional analysis of the fusion and attachment glycoproteins of mojiang
600 henipavirus. Viruses 13.
601 15. Lee B, Pernet O, Ahmed A a., Zeltina A, Beaty SM, Bowden T a. 2015. Molecular
602 recognition of human ephrinB2 cell surface receptor by an emergent African
603 henipavirus. Proceedings of the National Academy of Sciences 201501690.
604 16. Laing ED, Navaratnarajah CK, Cheliout S, Silva D, Petzing SR, Xu Y, Broder CC, Xu K.
605 2019. Structural and functional analyses reveal promiscuous and species specific
606 use of ephrin receptors by Cedar virus. Proceedings of the National Academy of
607 Sciences 116:20707–20715.
608 17. Rissanen I, Ahmed AA, Azarm K, Beaty S, Hong P, Nambulli S, Duprex WP, Lee B,
609 Bowden TA. 2017. Idiosyncratic Mòjiang virus attachment glycoprotein directs a

610 host-cell entry pathway distinct from genetically related henipaviruses. Nature
611 Communications 8.
612 18. Flick R, Walpita P, Czub M. 2006. Nipah and Hendra viral infections, p. 586–589. *In*
613 Tropical Infectious Diseases. Churchill Livingstone.
614 19. Rima B, Buschmann AB-, Dundon WG, Duprex P, Easton A, Fouchier R, Kurath G,
615 Lamb R, Lee B, Rota P, Wang L, Consortium IR. 2019. ICTV Virus Taxonomy Profile:
616 *Paramyxoviridae*. Journal of General Virology 100:1593–1594.
617 20. Calain P, Roux L. 1993. The rule of six, a basic feature for efficient replication of
618 Sendai virus defective interfering RNA. Journal of Virology 67:4822–4830.
619 21. Brook CE, Ranaivoson HC, Broder CC, Cunningham AA, Héraud J-M, Peel AJ, Gibson
620 L, Wood JLN, Metcalf CJ, Dobson AP. 2019. Disentangling serology to elucidate
621 henipa- and filovirus transmission in Madagascar fruit bats. Journal of Animal
622 Ecology 88:1001–1016.
623 22. Kettenburg G, Kistler A, Ranaivoson HC, Ahyong V, Andrianiaina A, Andry S, DeRisi
624 JL, Gentles A, Raharinosy V, Randriambolamanantsoa TH, Ravelomanantsoa NAF,
625 Tato CM, Dussart P, Heraud J-M, Brook CE. 2022. Full genome Nobecovirus
626 sequences from Malagasy fruit bats define a unique evolutionary history for this
627 coronavirus clade. Frontiers in Public Health 10.
628 23. Brook CE, Ranaivoson HC, Andriafidison D, Ralisata M, Razafimanahaka J, Héraud J,
629 Dobson AP, Metcalf CJ. 2019. Population trends for two Malagasy fruit bats.
630 Biological Conservation 234:165–171.
631 24. Ranaivoson HC, Héraud J-M, Goethert HK, Telford III SR, Rabetafika L, Brook CE.
632 2019. Babesial infection in the Madagascan flying fox, *Pteropus rufus* É. Geoffroy,
633 1803. Parasites & Vectors 1–13.
634 25. Brook CE, Bai Y, Dobson AP, Osikowicz LM, Ranaivoson HC, Zhu Q, Kosoy MY,
635 Dittmar K. 2015. Bartonella spp. in Fruit Bats and Blood-Feeding Ectoparasites in
636 Madagascar. PLoS Neglected Tropical Diseases 9.
637 26. Kalantar KL, Carvalho T, de Bourcy CFA, Dimitrov B, Dingle G, Egger R, Han J,
638 Holmes OB, Juan YF, King R, Kislyuk A, Lin MF, Mariano M, Morse T, Reynoso L v.,

639 Cruz DR, Sheu J, Tang J, Wang J, Zhang MA, Zhong E, Ahyong V, Lay S, Chea S, Bohl
640 JA, Manning JE, Tato CM, DeRisi JL. 2021. IDseq-An open source cloud-based
641 pipeline and analysis service for metagenomic pathogen detection and
642 monitoring. Gigascience 9:1–14.
643 27. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina
644 sequence data. Bioinformatics 30:2114–2120.
645 28. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M,
646 Gingeras TR. 2013. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics
647 29:15–21.
648 29. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nature
649 9:357–360.
650 30. Zhao Y, Tang H, Ye Y. 2012. RAPSearch2: A fast and memory-efficient protein
651 similarity search tool for next-generation sequencing data. Bioinformatics 28:125–
652 126.
653 31. Wu TD, Nacu S. 2010. Fast and SNP-tolerant detection of complex variants and
654 splicing in short reads. Bioinformatics 26:873–881.
655 32. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM,
656 Nikolenko SI, Pham S, Prjibelski AD, Pyshkin A v., Sirotkin A v., Vyahhi N, Tesler G,
657 Alekseyev MA, Pevzner PA. 2012. SPAdes: A new genome assembly algorithm and
658 its applications to single-cell sequencing. Journal of Computational Biology
659 19:455–477.
660 33. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment
661 search tool. Journal of Molecular Biology 215:403–410.
662 34. Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and
663 high throughput. Nucleic Acids Research 32:1792–1797.
664 35. Darriba Di, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. 2020.
665 ModelTest-NG: A new and scalable tool for the selection of DNA and protein
666 evolutionary models. Molecular Biology and Evolution 37:291–294.

667 36. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: A fast,
668 scalable and user-friendly tool for maximum likelihood phylogenetic inference.
669 Bioinformatics 35:4453–4455.
670 37. Felsenstein J. 1985. Confidence limits on phylogenies: An approach using the
671 bootstrap. Evolution (N Y) 39:783–791.
672 38. Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A. 2010.
673 How many bootstrap replicates are necessary? Journal of Computational Biology
674 17:337–354.
675 39. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut
676 A, Drummond AJ. 2014. BEAST 2: A software platform for Bayesian evolutionary
677 analysis. PLoS Computational Biology 10.
678 40. Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with
679 BEAUti and the BEAST 1.7. Molecular Biology and Evolution 29:1969–1973.
680 41. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior
681 summarization in Bayesian phylogenetics using Tracer 1.7. Systematic Biology
682 67:901–904.
683 42. Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by
684 sampling trees. BMC Evolutionary Biology 7.
685 43. Yu G, Smith DK, Zhu H, Guan Y, Lam TTY. 2017. Ggtree: an R Package for
686 visualization and annotation of phylogenetic trees with their covariates and other
687 associated data. Methods in Ecology and Evolution 8:28–36.
688 44. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O,
689 Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl
690 SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T,
691 Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M,
692 Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli
693 P, Hassabis D. 2021. Highly accurate protein structure prediction with AlphaFold.
694 Nature 596:583–589.

695 45. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH,
696 Ferrin TE. 2021. UCSF ChimeraX: Structure visualization for researchers, educators,
697 and developers. Protein Science 30:70–82.
698 46. Vanmechelen B, Bletsa M, Laenen L, Lopes AR, Vergote V, Beller L, Deboutte W,
699 Korva M, Avšič Županc T, Goüy de Bellocq J, Gryseels S, Leirs H, Lemey P, Vrancken
700 B, Maes P. 2018. Discovery and genome characterization of three new
701 Jeilongviruses, a lineage of paramyxoviruses characterized by their unique
702 membrane proteins. BMC Genomics 19.
703 47. Guillaume V, Aslan H, Ainouze M, Guerbois M, Fabian Wild T, Buckland R,
704 Langedijk JPM. 2006. Evidence of a potential receptor-binding site on the Nipah
705 virus G protein (NiV-G): Identification of globular head residues with a role in
706 fusion promotion and their localization on an NiV-G structural model. Journal of
707 Virology 80:7546–7554.
708 48. Negrete OA, Wolf MC, Aguilar HC, Enterlein S, Wang W, Mühlberger E, Su S v.,
709 Bertolotti-Ciarlet A, Flick R, Lee B. 2006. Two key residues in EphrinB3 are critical
710 for its use as an alternative receptor for Nipah virus. PLoS Pathogens 2:0078–0086.
711 49. Field H, Jordan D, Edson D, Morris S, Melville D, Parry-Jones K, Broos A, Divljan A,
712 McMichael L, Davis R, Kung N, Kirkland P, Smith C. 2015. Spatiotemporal Aspects
713 of Hendra Virus Infection in Pteropid Bats (Flying-Foxes) in Eastern Australia. Plos
714 One 10:e0144055.
715 50. Páez DJ, Giles J, Mccallum H, Field H, Jordan D, Peel AJ, Plowright RK. 2017.
716 Conditions affecting the timing and magnitude of Hendra virus shedding across
717 pteropodid bat populations in Australia. Epidemiology and Infection 145:3143–
718 3153.
719 51. Cappelle J, Furey N, Hoem T, Ou TP, Lim T, Hul V, Heng O, Chevalier V, Dussart P,
720 Duong V. 2021. Longitudinal monitoring in Cambodia suggests higher circulation of
721 alpha and betacoronaviruses in juvenile and immature bats of three species.
722 Scientific Reports 11:24145.

723 52. Hausmann S, Jacques J-P, Kolakofsky D. 1996. Paramyxovirus RNA editing and the
724 requirement for hexamer genome length. RNA 2:1033–1045.
725 53. Shi JJ, Chan LM, Peel AJ, Lai R, Yoder AD, Goodman SM. 2014. A deep divergence
726 time between sister species of eidolon (Pteropodidae) with evidence for
727 widespread panmixia. Acta Chiropterologica 16:279–292.
728 54. Douglas J, Drummond AJ, Kingston RL. 2021. Evolutionary history of
729 cotranscriptional editing in the paramyxoviral phosphoprotein gene. Virus
730 Evolution 7.
731 55. Satterfield BA, Cross RW, Fenton KA, Agans KN, Basler CF, Geisbert TW, Mire CE.
732 2015. The immunomodulating v and W proteins of Nipah virus determine disease
733 course. Nature Communications 6.
734 56. Patterson JB, Thomas D, Lewicki H, Billeter MA, Oldstone MBA. 2000. V and C
735 proteins of measles virus function as virulence factors in vivo. Virology 267:80–89.
736 57. Negrete OA, Chu D, Aguilar HC, Lee B. 2007. Single amino acid changes in the
737 Nipah and Hendra virus attachment glycoproteins distinguish ephrinB2 from
738 ephrinB3 usage. Journal of Virology 81:10804–10814.
739
740
741
742
743
744
745
746
747
748
749
750

751 **Table 1.** Prevalence of HNV infections in the urine of Malagasy bats captured during 2013-2019

752 collection period.

		By Site		By Season/Sex		
Roost site	**Species**	**Total sampled**	**Total henipavirus positive**	**Season**	**Total sampled # (M,F)**	**Total henipavirus positive # (M,F)**
Ambakoana	*P. rufus*	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	*E. dupreanum*	8	0	wet '17/'18	5 (1,4)	0 (0,0)
				dry '18	3 (1,2)	0 (0,0)
Angavokely	*E. dupreanum*	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	*P. rufus*	8	0	dry '14	8 (4,4)	0 (0,0)
Mahialambo	*P. rufus*	5	0	wet '18/'19	5 (2,3)	0 (0,0)
Makira	*P. rufus*	2	0	dry '15	2 (2,0)	0 (0,0)
Marovitsika	*P. rufus*	62	0	wet '17/'18	9 (3,6)	0 (0,0)
				dry '18	7 (3,4)	
				wet '18/'19	46 (20,26)	
Maromizaha	*R. madagascariensis*	5	0	wet '13	2 (1,1)	0 (0,0)
				dry '14	1 (0,1)	
				wet '14/'15	2 (2,0)	

753

754

755

756

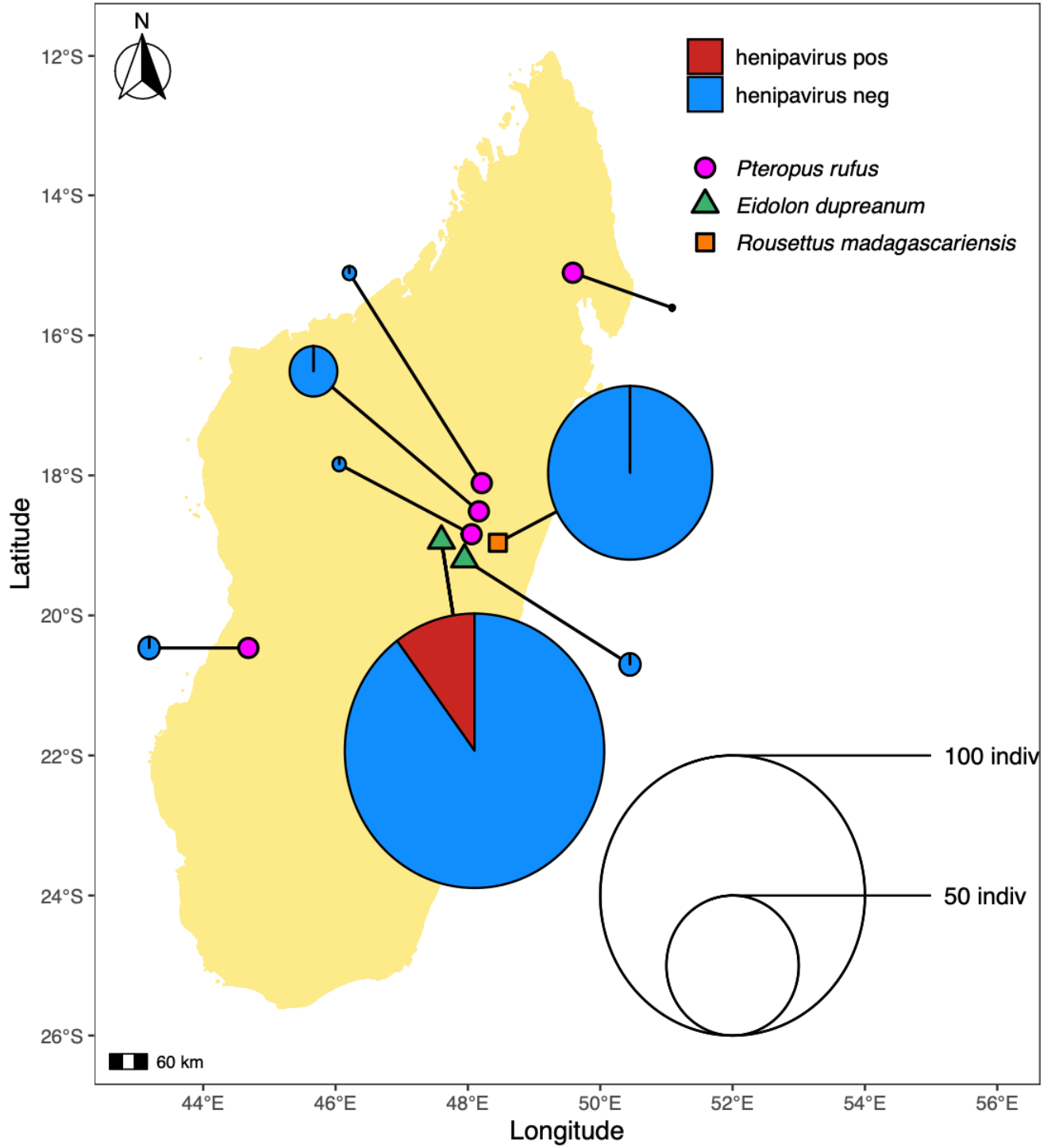
757

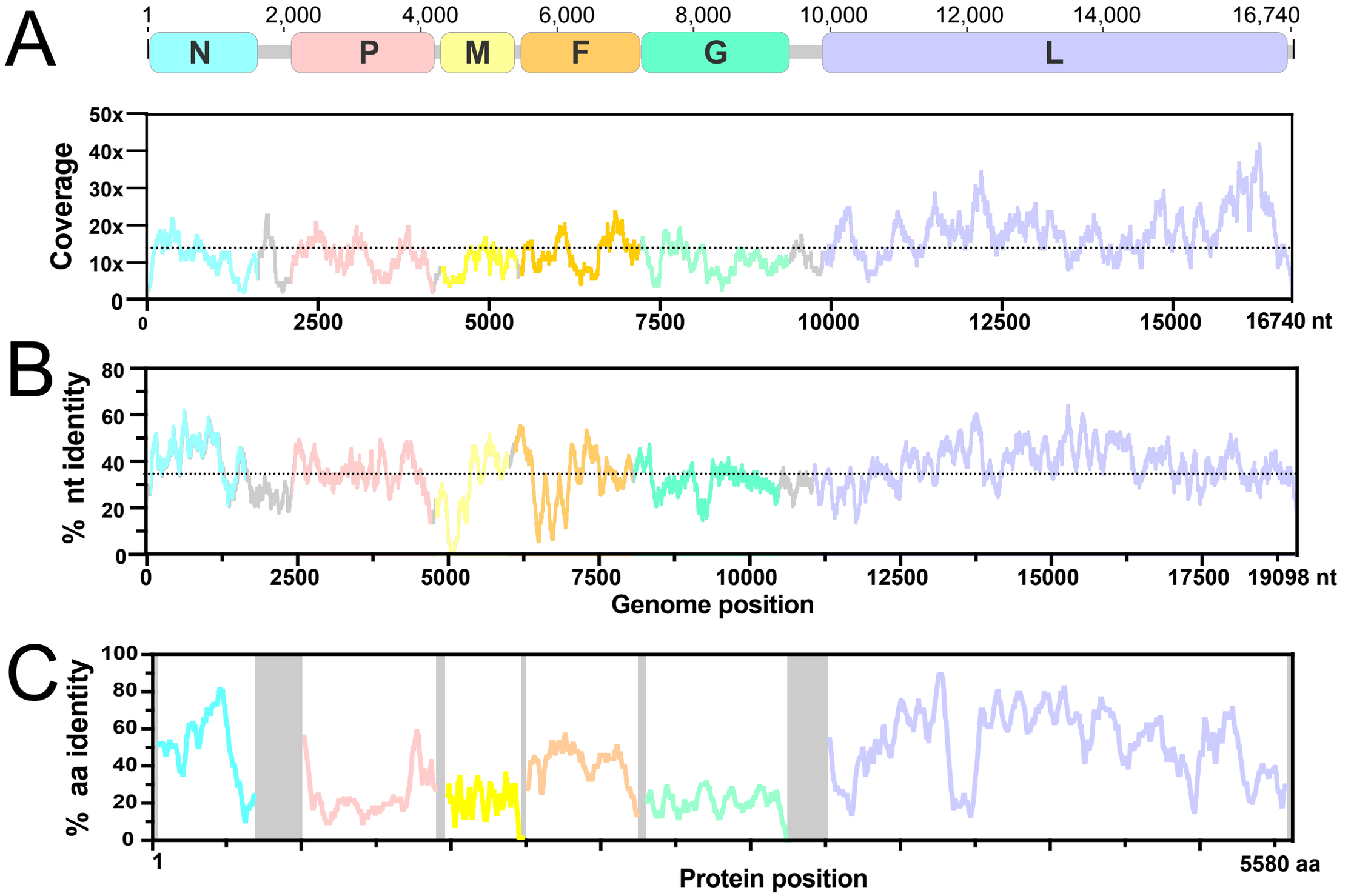
758

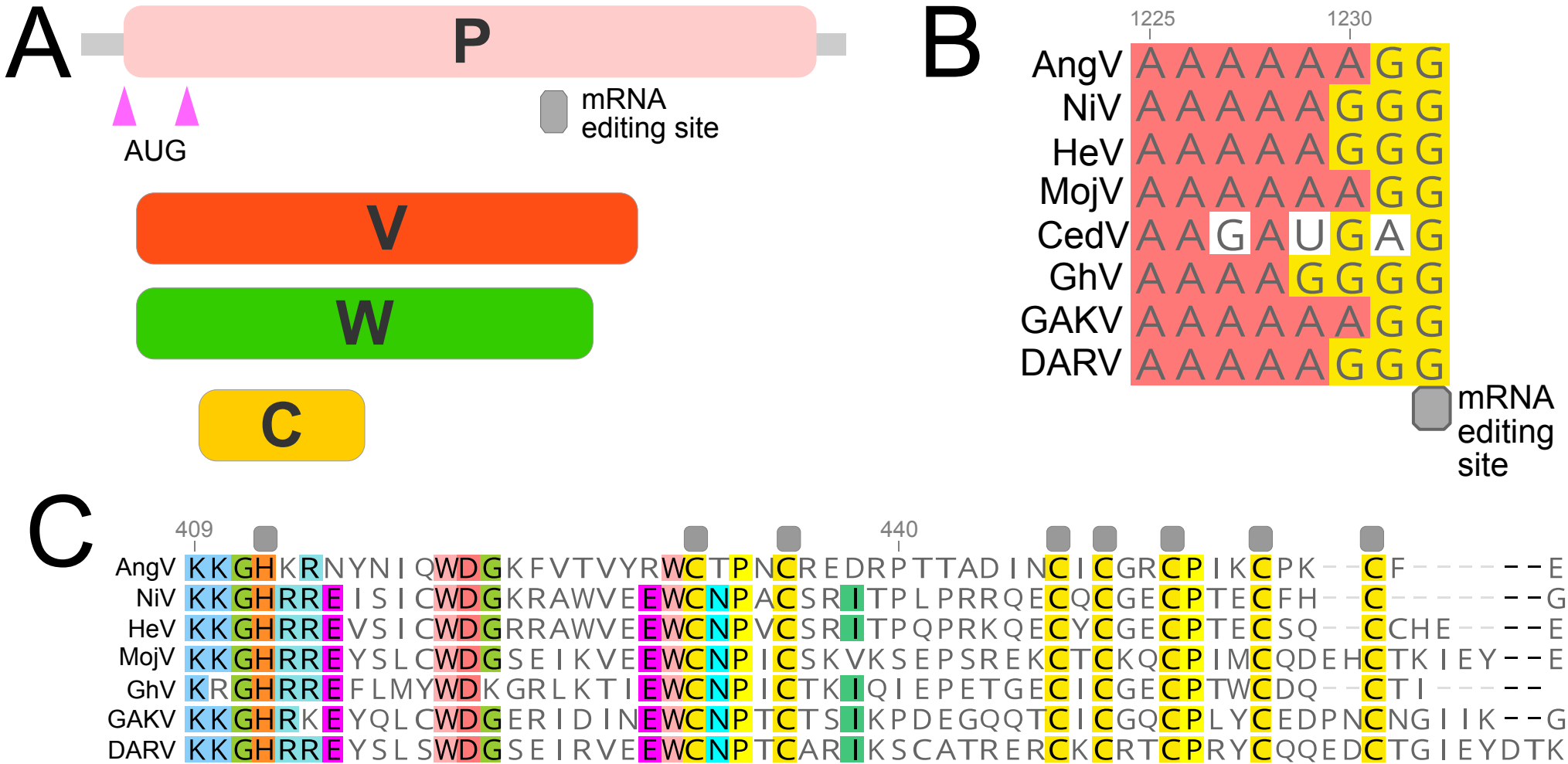
759 **Table 2.** Length and pairwise sequence identity of predicted open reading frames of AngV and
760 other HNV.
761

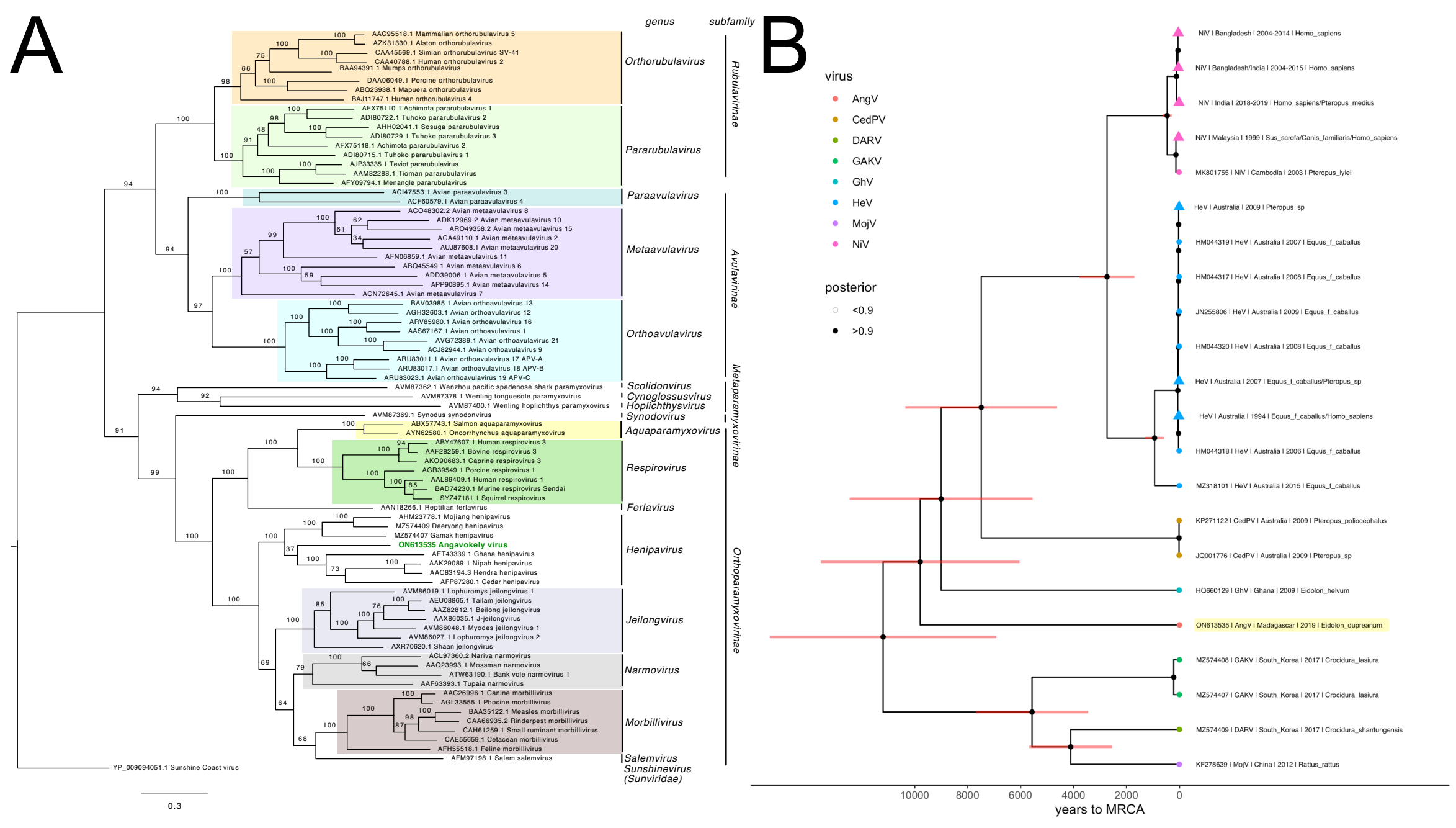
Gene	AngV *length aa*	NiV *length aa (%nt, %aa)*	HeV	MojV	CedV	GAKV	DARV	GhV
N	514	532 (56.4, 48.0)	532 (55.4, 46.7)	539 (55.8, 45.3)	510 (57.5, 48.3)	533 (58.0, 47.5)	574 (55.3, 24.0)	514 (56.2, 48.9)
P	693	709 (47.8, 25.1)	707 (48.1, 24.1)	694 (47.6, 23.6)	737 (42.3, 21.6)	586 (43.8, 22.5)	698 (47.3, 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)

762
763
764
765
766
767


N
henipavirus pos
henipavirus neg
Pteropus rufus
Eidolon dupreanum
Rousettus madagascariensis
100 indiv
50 indiv
60 km
12°S
14°S
16°S
18°S
20°S
22°S
24°S
26°S
Latitude
44°E
46°E
48°E
50°E
52°E
54°E
56°E
Longitude


A
1
2,000
4,000
6,000
8,000
10,000
12,000
14,000
16,740
N
P
M
F
G
L
Coverage
50x
40x
30x
20x
10x
0
0
2500
5000
7500
10000
12500
15000
16740 nt
B
% nt identity
80
60
40
20
0
0
2500
5000
7500
10000
12500
15000
17500
19098 nt
Genome position
C
% aa identity
100
80
60
40
20
0
1
5580 aa
Protein position


A
P
AUG
mRNA editing site
V
W
C
B
1225
1230
AngV AAAAAAGG
NiV AAAAAGGG
HeV AAAAAGGG
MojV AAAAAAGG
CedV AAGAUGAG
GhV AAAAGGGG
GAKV AAAAAAGG
DARV AAAAAGGG
mRNA editing site
C
409
440
AngV KKGHKRNYNIQWDGKFVTVYRWCTPNCREDRPTTADINCICGRCPIKCPK--CF----E
NiV KKGHRREISICWDGKRAWVEEWCNPACSRITPLPRRQECQCGECPTECFH--C----G
HeV KKGHRREVSICWDGRRAWVEEWCNPVCSRITPQPRKQECYCGECPTECSQ--CCHE--E
MojV KKGHRREYSLCWDGSEIKVEEWCNPICSKVKSEPSREKCTCKQCPIMCQDEHCTKIEY--E
GhV KRGHRREFLMYWDKGRLKTIEWCNPICTKIQIEPETGECICGECPTWCDQ--CTI---
GAKV KKGHRKEYQLCWDGERIDINEWCNPTCTSIKPDEGQQTCICGQCPLYCEDPNCNGIIK--G
DARV KKGHRREYSLSWDGSEIRVEEWCNPTCARIKSCATRERCKCRTCPRYCQQEDCTGIEYDTK


A
genus
subfamily
AAC95518.1 Mammalian orthorubulavirus 5
AZK31330.1 Alston orthorubulavirus
CAA45569.1 Simian orthorubulavirus SV-41
CAA40788.1 Human orthorubulavirus 2
BAA94391.1 Mumps orthorubulavirus
DAA06049.1 Porcine orthorubulavirus
ABQ23938.1 Mapuera orthorubulavirus
BAJ11747.1 Human orthorubulavirus 4
Orthorubulavirus
AFX75110.1 Achimota pararubulavirus 1
ADI80722.1 Tuhoko pararubulavirus 2
AHH02041.1 Sosuga pararubulavirus
ADI80729.1 Tuhoko pararubulavirus 3
AFX75118.1 Achimota pararubulavirus 2
ADI80715.1 Tuhoko pararubulavirus 1
AJP33335.1 Teviot pararubulavirus
AAM82288.1 Tioman pararubulavirus
AFY09794.1 Menangle pararubulavirus
Pararubulavirus
Rubulavirinae
ACI47553.1 Avian paraavulavirus 3
ACF60579.1 Avian paraavulavirus 4
Paraavulavirus
ACO48302.2 Avian metaavulavirus 8
ADK12969.2 Avian metaavulavirus 10
ARO49358.2 Avian metaavulavirus 15
ACA49110.1 Avian metaavulavirus 2
AUJ87608.1 Avian metaavulavirus 20
AFN06859.1 Avian metaavulavirus 11
ABQ45549.1 Avian metaavulavirus 6
ADD39006.1 Avian metaavulavirus 5
APP90895.1 Avian metaavulavirus 14
ACN72645.1 Avian metaavulavirus 7
Metaavulavirus
BAV03985.1 Avian orthoavulavirus 13
AGH32603.1 Avian orthoavulavirus 12
ARV85980.1 Avian orthoavulavirus 16
AAS67167.1 Avian orthoavulavirus 1
AVG72389.1 Avian orthoavulavirus 21
ACJ82944.1 Avian orthoavulavirus 9
ARU83011.1 Avian orthoavulavirus 17 APV-A
ARU83017.1 Avian orthoavulavirus 18 APV-B
ARU83023.1 Avian orthoavulavirus 19 APV-C
Orthoavulavirus
Avulavirinae
AVM87362.1 Wenzhou pacific spadenose shark paramyxovirus
AVM87378.1 Wenling tonguesole paramyxovirus
AVM87400.1 Wenling hoplichthys paramyxovirus
Scolidonvirus
Cynoglossusvirus
Hoplichthysvirus
AVM87369.1 Synodus synodonvirus
Synodovirus
ABX57743.1 Salmon aquaparamyxovirus
AYN62580.1 Oncorrhynchus aquaparamyxovirus
Aquaparamyxovirus
ABY47607.1 Human respirovirus 3
AAF28259.1 Bovine respirovirus 3
AKO90683.1 Caprine respirovirus 3
AGR39549.1 Porcine respirovirus 1
AAL89409.1 Human respirovirus 1
BAD74230.1 Murine respirovirus Sendai
SYZ47181.1 Squirrel respirovirus
Respirovirus
Metaparamyxovirinae
AAN18266.1 Reptilian ferlavirus
Ferlavirus
AHM23778.1 Mojiang henipavirus
MZ574409 Daeryong henipavirus
MZ574407 Gamak henipavirus
ON613535 Angavokely virus
AET43339.1 Ghana henipavirus
AAK29089.1 Nipah henipavirus
AAC83194.3 Hendra henipavirus
AFP87280.1 Cedar henipavirus
Henipavirus
AVM86019.1 Lophuromys jeilongvirus 1
AEU08865.1 Tailam jeilongvirus
AAZ82812.1 Beilong jeilongvirus
AAX86035.1 J-jeilongvirus
AVM86048.1 Myodes jeilongvirus 1
AVM86027.1 Lophuromys jeilongvirus 2
AXR70620.1 Shaan jeilongvirus
Jeilongvirus
ACL97360.2 Nariva narmovirus
AAQ23993.1 Mossman narmovirus
ATW63190.1 Bank vole narmovirus 1
AAF63393.1 Tupaia narmovirus
Narmovirus
AAC26996.1 Canine morbillivirus
AGL33555.1 Phocine morbillivirus
BAA35122.1 Measles morbillivirus
CAA66935.2 Rinderpest morbillivirus
CAH61259.1 Small ruminant morbillivirus
CAE55659.1 Cetacean morbillivirus
AFH55518.1 Feline morbillivirus
Morbillivirus
AFM97198.1 Salem salemvirus
Salemvirus
Orthoparamyxovirinae
YP_009094051.1 Sunshine Coast virus
Sunshinevirus
(Sunviridae)
0.3
B
virus
AngV
CedPV
DARV
GAKV
GhV
HeV
MojV
NiV
posterior
<0.9
>0.9
NiV | Bangladesh | 2004-2014 | Homo_sapiens
NiV | Bangladesh/India | 2004-2015 | Homo_sapiens
NiV | India | 2018-2019 | Homo_sapiens/Pteropus_medius
NiV | Malaysia | 1999 | Sus_scrofa/Canis_familiaris/Homo_sapiens
MK801755 | NiV | Cambodia | 2003 | Pteropus_lylei
HeV | Australia | 2009 | Pteropus_sp
HM044319 | HeV | Australia | 2007 | Equus_f_caballus
HM044317 | HeV | Australia | 2008 | Equus_f_caballus
JN255806 | HeV | Australia | 2009 | Equus_f_caballus
HM044320 | HeV | Australia | 2008 | Equus_f_caballus
HeV | Australia | 2007 | Equus_f_caballus/Pteropus_sp
HeV | Australia | 1994 | Equus_f_caballus/Homo_sapiens
HM044318 | HeV | Australia | 2006 | Equus_f_caballus
MZ318101 | HeV | Australia | 2015 | Equus_f_caballus
KP271122 | CedPV | Australia | 2009 | Pteropus_poliocephalus
JQ001776 | CedPV | Australia | 2009 | Pteropus_sp
HQ660129 | GhV | Ghana | 2009 | Eidolon_helvum
ON613535 | AngV | Madagascar | 2019 | Eidolon_dupreanum
MZ574408 | GAKV | South_Korea | 2017 | Crocidura_lasiura
MZ574407 | GAKV | South_Korea | 2017 | Crocidura_lasiura
MZ574409 | DARV | South_Korea | 2017 | Crocidura_shantungensis
KF278639 | MojV | China | 2012 | Rattus_rattus
10000
8000
6000
4000
2000
0
years to MRCA

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