Discovery of Flaviviridae-derived endogenous viral elements in shrew genomes provide novel insights into *Pestivirus* ancient history

Li YQ[1], Bletsa M[1], Zisi Z[1], Boonen I[1], Gryseels S[1,2], Kafetzopoulou L[1], Webster J[3], Pybus OG[3], Van de Perre F[2], Li HT[4], Li YY[4], Li YC[4], Abramov A[5], Lymberakis P[6], Lemey P[1], Lequime S[1,7*]

1 Department of Microbiology, Immunology and Transplantation, KU Leuven, Rega Institute, KU Leuven, Leuven, Belgium
2 Evolutionary Ecology Group, University of Antwerp, Antwerp, Belgium
3 Department of Pathobiology and Population Science, Royal Veterinary College, University of London, London, UK
4 Marine College, Shandong University (Weihai), Weihai, China
5 Laboratory of Theriology, Zoological Institute of the Russian Academy of Sciences, Saint Petersburg, Russia
6 Natural History Museum of Crete, Greece
7 Cluster of Microbial Ecology, Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, The Netherlands.

* corresponding author: Sebastian Lequime (s.j.j.lequime@rug.nl)

Keywords: Endogenous viral element, pestivirus, *Flaviviridae*, *Crocidura*, host range, paleovirology

Abstract

As viral genomic imprints in host genomes, endogenous viral elements (EVEs) shed light on the deep evolutionary history of viruses, ancestral host ranges, and ancient viral-host interactions. In addition, they may provide crucial information for calibrating viral evolutionary timescales. In this study, we conducted a comprehensive *in silico* screening of a large dataset of available mammalian genomes for EVEs deriving from members of the viral family *Flaviviridae*, an important group of viruses including well-known human pathogens. We identified two novel pestivirus-like EVEs in the reference genome of the Indochinese shrew (*Crocidura indochinensis*). Homologs of these novel EVEs were subsequently detected *in vivo* by molecular detection and sequencing in 27 shrew species, including 26 species representing a wide distribution within the Crocidurinae subfamily and one in the Soricinae subfamily. Based on this wide distribution, we estimate that the integration event occurred before the last common ancestor of the subfamily, about 10.8 million years ago, attesting to an ancient origin of pestiviruses and *Flaviviridae* in general. Moreover, we provide the first description of *Flaviviridae*-derived EVEs in mammals even though the family encompasses numerous mammal-infecting members, including major human pathogens such as Zika, dengue, or hepatitis C viruses. This also suggests that shrews were past and perhaps also current natural reservoirs of pestiviruses. Taken together, our results expand the current known *Pestivirus* host range and provide novel insight into the ancient evolutionary history of pestiviruses and the *Flaviviridae* family in general.

1 – Introduction

Endogenous viral elements (EVEs) are integrations of partial or full-length viral genomic material into the host genome (Katzourakis and Gifford, 2010a). In addition to retroviruses, which incorporate their genomic sequences into their host genome as an essential part of their replication cycle, many eukaryotic viruses have been found endogenized in various hosts (Feschotte and Gilbert, 2012). These non-retroviruses can derive from dsDNA (Aswad and Katzourakis, 2014; Li and Li, 2015; Liu et al., 2020), ssDNA (Belyi et al., 2010a; Katzourakis and Gifford, 2010a; Kobayashi et al., 2019), dsRNA (Horie et al., 2010; Katzourakis and Gifford, 2010b; Liu et al., 2012), and even ssRNA viruses (Crochu et al., 2004; Flynn and Moreau, 2019; Horie et al., 2010; Lequime and Lambrechts, 2017). Non-retroviral RNA virus-derived EVEs arise from a conjunction of relatively rare events: i) production of DNA genomic material, using retrotransposon encoded reverse transcriptase (Horie, 2019; Horie et al., 2010), ii) integration in the host chromosome of germ-line cells, and iii) overcoming genetic drift and/or natural selection at the population until fixation (Aiewsakun and Katzourakis, 2015a; Holmes, 2011). EVEs thus reflect long-term and intimate interactions of viruses with their hosts, and their identification can reveal insights into past and present host distributions of viral genera and families. The detection of endogenous bornavirus-like elements in invertebrate genomes (Horie et al., 2013), for example, suggested a broader host range for bornaviruses than previously thought. The study of filovirus-like EVEs in some small mammals offer predictive value for further identifying filovirus reservoirs (Taylor et al., 2010). Similarly, the discovery of flavivirus-derived EVEs in *Anopheles* mosquito genomes supported the idea that *Anopheles* mosquitoes could also be the natural hosts of flaviviruses (Lequime and Lambrechts, 2017), as confirmed by other studies (Colmant et al., 2017; Öncü et al., 2018).

Aside from qualitative insights into host ranges of viruses, EVEs can also shed light on deep evolutionary histories of viruses. EVEs are significant traces of past virus-host interactions; unlike animals or plants, viruses do not leave physical fossil records, limiting our ability to study their deep evolutionary histories. EVEs could thus be considered as “genomic fossil records” that can help to unravel long-term evolutionary dynamics between hosts and viral families. The presence of EVE homologs in different host species hints at an integration event before speciation, here the time to the most recent common ancestor. It thus provides a minimum age estimate for the integration event, and therefore a minimum age for the existence of a specific viral taxonomic group (Aiewsakun and Katzourakis, 2015b). For example, EVEs derived from adeno-associated viruses appear to be orthologous in African and Asian elephants,

78 indicating an integration event more than 6 million years ago (Kobayashi et al., 2019).
79 Similarly, the discovery of abundant bornavirus-like EVEs across vertebrates reveals that
80 ancient bornaviral infections occurred over a timeframe of about 100 million years before
81 present (Kawasaki et al., 2021). These studies can also provide genetic fossil calibration points
82 for further gauging ancient viral timescales using phylogenetics (Feschotte and Gilbert, 2012).
83
84 Members of the *Flaviviridae* family are linear, positive-sense, single-stranded RNA viruses
85 currently classified in four recognized genera (*Flavivirus*, *Hepacivirus*, *Pestivirus*, and
86 *Pegivirus*). They encompass many significant pathogens, such as dengue, Zika, hepatitis C, or
87 bovine viral diarrhea virus. In addition to human and livestock infections, *Flaviviridae* viruses
88 have been detected in a broad range of hosts, e.g. non-human primates (Mares-Guia et al.,
89 2020), rodents (Bletsa et al., 2021), bats (Wu et al., 2018), birds (Strand et al., 2018), arthropods
90 (Shi et al., 2016), and fish (Hartlage, 2016; Soto et al., 2020). Divergence date estimates on a
91 genus-wide scale are relatively rare, with the oldest, for the *Flavivirus* genus, being estimated
92 at 200,000 years ago (Pettersson and Fiz-Palacios, 2014). Although *Flaviviridae* viruses have
93 been detected in various hosts, especially in small mammals (Bletsa et al., 2021; de Lamballerie
94 et al., 2002; Wu et al., 2020, 2018), current published studies have only identified *Flaviviridae*-
95 derived EVEs in arthropods, including mosquitos (Crochu et al., 2004; Lequime and
96 Lambrechts, 2017; Roiz et al., 2009), ticks (Maruyama et al., 2014), and crustaceans (Parry
97 and Asgari, 2019), and these are all related to the *Flavivirus* genus. A recent study also
98 identified *Flaviviridae*-derived EVEs in a wide variety of hosts, including various invertebrates
99 and fish (Bamford et al., 2021). Currently, however, convincing evidence for *Flaviviridae*-
100 derived EVEs in vertebrates remains lacking. Interestingly, potential integration has been
101 suggested in medaka fish (Belyi et al., 2010b) as well as in rabbit and hare genomes (Silva et
102 al., 2015, 2012). While this raises the hypothesis that *Flaviviridae* viruses have integrated in
103 vertebrate hosts, the origin of these specific genomic sequences remains inconclusive due to
104 their short size and low sequence similarity to the *Flaviviridae* (*Flavivirus* and *Hepacivirus*
105 respectively).
106
107 In this study, we explored *in silico* the presence of *Flaviviridae*-derived EVEs in a
108 comprehensive set of mammalian genomes, and we discovered two novel pesti-like EVEs in
109 the genome of the Indochinese shrew *Crocidura indochinensis.* We subsequently identified
110 and characterized homologs of these EVEs *in vivo* in 26 species of the Crocidurinae subfamily
111 and one member of the Soricinae subfamily, establishing the integration event at least 10.8

112 million years ago. Our results provide the first evidence for an ancient origin of pestiviruses
113 and also contribute to a better understanding of the evolutionary history of the *Flaviviridae*
114 family in general.
115
116 **2 – Material and Methods**
117 **2.1 – *In silico* survey**
118 **2.1.1 – Data collection**
119 To screen for *Flaviviridae*-like EVEs, 689 mammalian genomes (57 bats, 9 insectivores, 177
120 rodents, 101 nonhuman primates, 207 even-toed ungulates, 15 odd-toed ungulates, 108
121 carnivores, and 15 marsupials), were retrieved from the National Center for Biotechnology
122 Information (NCBI) Whole Genome Shotgun (WGS) database (last accessed in November
123 2020). A detailed list of all the surveyed mammalian genomes is provided in Supplementary
124 Table S1. A representative group of 306 *Flaviviridae* or *Flaviviridae*-like polyprotein
125 sequences was compiled from the NCBI non-redundant protein database (accessed in February
126 2019). We provide a list of the nucleotide/protein accession numbers in Supplementary Table
127 S2.
128
129 **2.1.2 – Genome screening**
130 *Flaviviridae* polyprotein sequences were used as queries in tBLASTn (BLAST+ v2.6.0)
131 (Camacho et al., 2009) searches with mammalian genomes as targets. To avoid potential
132 artifacts, only hits with E-value < 10^{-4} and length >= 250 nt were extracted from mammalian
133 genomes based on the reported position by BLAST in the host contig. These putative EVEs
134 were then used as query in a reciprocal tBLASTx (BLAST+ v2.6.0) (Camacho et al., 2009)
135 against a local NCBI nucleotide (nt) database (accessed in October 2018) and BLASTx
136 (BLAST+ v2.6.0) (Camacho et al., 2009) against a non-redundant protein (nr) database
137 (accessed in October 2018). EVEs were confirmed if the best hits contained *Flaviviridae* family
138 members with an E-value < 10^{-4} and length >= 250 nt. The presence of conserved viral genetic
139 features within the hits was assessed using the NCBI Conserved Domain Database (Marchler-
140 Bauer et al., 2015).
141
142 **2.1.3 – EVE characterization**
143 Upon identification of the EVEs, they were translated and aligned with corresponding
144 polyprotein sequences from several representative *Flaviviridae* species using MAFFT v7.453
145 (Katoh et al., 2002). All alignments were trimmed in BMGE v1.12 (Criscuolo and Gribaldo,

146 2010) in order to select for phylogenetic informative regions. The best substitution models
147 were PMB+G4 for the EVE1 alignment, LG+F+G4 for the EVE2 alignment, and LG+F+I+G4
148 for the concatenated EVEs alignment according to the BIC criterion and were used to construct
149 phylogenetic ML trees with IQ-TREE v1.6.12 (Nguyen et al., 2015).
150
151 **2.1.4 – Flanking region analysis**
152 To characterize the EVEs loci and identify potential transposable elements or other genetic
153 features, flanking regions of the identified EVEs were extracted from the host contigs and used
154 as BLAST queries to screen against the NCBI nucleotide (nt) and non-redundant protein (nr)
155 databases (both accessed in October 2018).
156
157 **2.1.5 – Metagenomic screening**
158 According to the WGS screening results above, some *Flaviviridae*-related hits were detected
159 in a shrew (*Crocidura indochinensis*) genome. However, apart from the *Crocidura*
160 *indochinensis* and *Sorex araneus* complete genomes, only a limited number of shrew genomes
161 are currently available in the NCBI WGS database. Therefore, 73 DNA experimental genomic
162 data sets and 6 RNA-Seq transcriptome data sets (Supplementary Table S3) from the Soricidae
163 family were retrieved from NCBI Sequence Read Archive (SRA) database using SRA Toolkit
164 v2.10.8 (Leinonen et al., 2011). Reads were mapped to the identified EVEs nucleotide
165 references using Bowtie2 v2.3.5.1 (Langmead and Salzberg, 2012). Alignment files were
166 processed with SAMtools v1.10 (Li et al., 2009) and coverage was determined using bedtools
167 v2.27.1 (Quinlan and Hall, 2010) and visualized in RStudio v1.1.463.
168
169 **2.2 – *In vivo* validation**
170 **2.2.1 – Sample collection**
171 Based on the screening results, to further verify the presence of *Flaviviridae*-related EVEs *in*
172 *vivo*, a total of 65 tissue and DNA samples from species belonging to the *Crocidura* genus and
173 6 other related genera of the Soricidae family, namely *Paracrocidura*, *Scutisorex*, *Suncus*,
174 *Sylvisorex* (subfamily Crocidurinae), *Neomys,* and *Sorex* (subfamily Soricinae), were screened
175 for the presence of the identified EVEs. These samples were previously collected in China,
176 Vietnam, Africa, and the Eastern Mediterranean (Supplementary Table S4) as part of other
177 studies (Bannikova et al., 2011; de Perre et al., 2019; Jenkins et al., 2013; van de Perre et al.,
178 2018)
179

180 **2.2.2 – Target EVEs and cytochrome b (Cytb) amplification**

181 DNA was extracted from tissue samples using the DNeasy Blood & Tissue Kit (Qiagen)
182 following the manufacturer's instructions.

183 To screen for the presence of EVEs *in vivo*, we designed 18 PCR primers (Supplementary
184 Table S5) spanning the 2 EVEs region and a section of the intermediate flanking region from
185 the host genome. Amplicons were generated with DreamTaq DNA Polymerase (ThermoFisher
186 Scientific) using the following cycling conditions: (i) 3 min of denaturation at 95°C, (ii) 35
187 cycles of 95°C for 30s, 56°C for 30s, 72°C extension for 1min/kb, and (iii) 10 min final
188 extension at 72°C.

189 To confirm the host species, and to complement available specimen information, the
190 mitochondrial Cytb gene was amplified using general primers (Supplementary Table S5) of the
191 Crocidurinae subfamily. PCR reactions were conducted using the DreamTaq DNA Polymerase
192 (ThermoFisher Scientific) with the following thermal cycling conditions: (i) 3 min of
193 denaturation at 95°C, (ii) 35 cycles of 95°C for 30s, 56°C for 30s, 72°C extension for 1min/kb,
194 and (iii) 10 min final extension at 72°C.

195 All PCR products were purified using the ExoSAP-IT PCR Product Cleanup (ThermoFisher
196 Scientific) or Zymoclean Gel DNA Recovery Kits (ZYMO Research) to remove primer dimers
197 and unspecific products, following the manufacturer's instructions.

198

199 **2.2.3 – Sanger sequencing**

200 The generated PCR products were sequenced by Macrogen Europe. The amplicons were
201 mapped to the whole EVEs region (2,235nt) from the WGS *Crocidura indochinensis* contig,
202 and concatenated based on consensus sequence to get the complete EVEs in Geneious Prime®
203 v2020.2.4. Cytb amplicons (~1,140nt) were forward and reverse sequenced and a consensus
204 sequence was generated using Geneious Prime® v2020.2.4.

205

206 **2.2.4 – MinION sequencing**

207 For 12 samples with relatively low-quality Sanger sequencing chromatograms (additional
208 information provided in Supplementary Table S6), MinION sequencing was performed to
209 obtain the complete EVEs region (~2,235nt) together with the Cytb gene (~1,140nt). The
210 Oxford Nanopore Technologies (ONT) 1D Native barcoding genomic DNA protocol was used
211 without the DNA fragmentation step and the barcoded amplicons were loaded onto the
212 MinION device. We used the MinKNOW software v19.13.5 on the MinIT companion for data

213 acquisition and basecalling. Qcat v1.1.0 (ONT, https://github.com/nanoporetech/qcat) was
214 used to demultiplex reads under the epi2me algorithm and to trim bad quality reads and
215 adapters with min score of 90. The EVE regions extracted from the WGS *Crocidura*
216 *indochinensis* contig were used as references to map the reads with Minimap2 v.2.22 (Li, 2018)
217 using -ax map-ont parameters. Alignments were converted and indexed using SAMtools v1.10
218 (Li et al., 2009) and consensus sequences were generated using a custom Python script
219 (Kafetzopoulou, 2019).
220
221 **2.3 – Phylogenetic analysis and visualization**
222 All generated EVEs sequences were translated and aligned with homologous polyproteins from
223 available *Pestivirus* species using MAFFT v7.453 (Katoh et al., 2002). Sequences of dengue-
224 2 and Zika virus (*Flavivirus* genus) were used as an outgroup. The alignment was trimmed
225 using BMGE v1.12 (Criscuolo and Gribaldo, 2010) and the filtered regions were used to
226 construct maximum-likelihood (ML) phylogenetic trees using IQ-TREE v1.6.12 (Nguyen et
227 al., 2015) under the best-fitting models (according to the Bayesian information criterion):
228 WAG+G4 (for EVE1), LG+F+I+G4 (for EVE2), and LG+G4 (for complete concatenated
229 EVEs). Phylogenies were visualized and annotated using FigTree v1.4.4 (A. Rambaut;
230 http://tree.bio.ed.ac.uk/software/figtree/). Percent identity matrices were generated using
231 Clustal Omega (Sievers et al., 2011) via EMBL-EBI web services (Madeira et al., 2019).
232 Cytb sequences of EVEs-positive specimens (n=48) were aligned using MAFFT v7.453 (Katoh
233 et al., 2002) together with a dataset of *n*=393 Soricidae nucleotide sequences downloaded from
234 NCBI. The generated alignment was trimmed in BMGE v1.12 (Criscuolo and Gribaldo, 2010)
235 and an ML phylogeny was reconstructed using IQ-TREE v1.6.12 (Nguyen et al., 2015) with
236 the best-fitting model (TIM2+F+I+G4).
237 To highlight the evolutionary relationships of our newly discovered EVEs and their hosts,
238 EVEs and Cytb phylogenies were annotated in ggtree v1.14.6 (Yu et al., 2017) and treeio v1.6.2
239 (Wang et al., 2020) R packages, followed by the estimation of a co-phylogenetic plot
240 (tanglegram) using the ape v5.0 (Paradis and Schliep, 2019) and dendextend v1.14.0 (Galili,
241 2015) R packages.
242
243 **2.4 – Characterization of selective pressure**
244 We only characterized the selective pressure acting on the EVE 1 locus, as the EVE 2 locus
245 exhibits widespread stop codons and translation frame shifts. We aligned the open reading
246 frame of the complete EVE1 region (318 nt) using MEGA v11.0.9 (Tamura et al., 2021). An

247 ML tree was build based on this alignment in IQ-TREE v1.6.12 (Nguyen et al., 2015) under
248 the best-fitting model (HKY+F). We then conducted two site-specific selection analyses to
249 characterize the selective pressure on each site using estimates of the ratio of non-
250 synonymous/synonymous substitution rate ($\omega = d_N/d_S$): 1) fixed effects likelihood (FEL)
251 analyses (Kosakovsky Pond and Frost, 2005) using MG94xREV model available in HyPhy
252 software v2.5.3 (Kosakovsky Pond et al., 2005), and 2) the Bayesian renaissance counting
253 method (Lemey et al., 2012) implemented in BEAST v1.10.5 (Suchard et al., 2018). The value
254 of ω quantifies the selective pressure, with $\omega > 1$ suggesting positive selection, $\omega = 1$ neutral
255 evolution and $\omega < 1$ negative or purifying selection.
256
257 **3 – Results**
258 **3.1 – False positive Flaviviridae-like hits from mammalian genome screening**
259 Our initial screening of 689 available mammalian genomes using *Flaviviridae* and
260 *Flaviviridae*-like polyproteins yielded 66 positive hits from 49 species, including rodents, non-
261 human primates, marsupials, insectivores, carnivores, and bats. Detailed information about our
262 *in silico* screening results can be found in Supplementary Table S7. All hits were similar to
263 three pestiviruses, namely border disease virus, bovine viral diarrhea virus 1, and bovine viral
264 diarrhea virus 2. With the exception of one shrew species (see section below), the position of
265 all hits in the corresponding viral genomic sequence was in the ubiquitin-homolog domain
266 between the nonstructural proteins NS2 and NS3, while some of the hits slightly expanded the
267 alignment to the NS3 region (shown in Supplementary Fig. 1). The ubiquitin domain in bovine
268 viral diarrhea virus is however predicted to originate from cellular derived insertions
269 in cytopathogenic pestivirus (Agapov et al., 2004; Becher and Tautz, 2011). The similarity
270 between this viral genomic region and ubiquitin poses a considerable risk for false positives
271 when searching for pestivirus-derived EVEs, and these hits were therefore not further
272 considered.
273
274 **3.2 – *In silico* identification of *Crocidura indochinensis* pesti-like EVEs**
275 Besides the ubiquitin-related false positive results, our *in silico* screening identified a series of
276 five *Flaviviridae*-related EVEs fragments in a single contig of the *Crocidura indochinensis*
277 reference genome PVKC01 (Table 1). The first EVE (EVE1) is 318 nt long, with its closest
278 BLAST hit being the Linda virus (Pestivirus) envelope glycoprotein E2 region (tBLASTx, 25.5%
279 identity, e-value $3.31E^{-35}$), without any stop codon (Fig. 1). The remaining four EVEs

280 fragments are 1,053 nt, 84 nt, 114 nt and 87 nt long respectively, with their closest BLAST hits
281 being a classical swine fever virus and a rodent pestivirus (with minimum amino-acid identity
282 24.9%, maximum 62.1%). These four fragments are separated by very short gaps, with lengths
283 of 16, 1 and 4 nucleotides, respectively. The two central fragments of EVE2, fragments 3 (84
284 nt) and fragment 4 (114 nt), are in a different translation frame than the two others, but in the
285 same orientation. Their arrangement in the host contig reflects their relative position in the
286 pestiviral genome, which partially spans the non-structural NS2 and NS3 genes (Fig 1). For
287 these reasons, in our further analyses, we considered these four fragments as the result of a
288 single pestiviral integration event, and thus a unique EVE (EVE2). Phylogenetic
289 reconstructions of the identified *Crocidura indochinensis* EVE1, EVE2 and entire
290 concatenated sequence with exogenous *Flaviviridae* viruses supports the pestivirus-origin of
291 the EVEs (Supplementary Fig. 2).
292 In addition, a fragment prior to EVE1 in the contig shows a strong similarity (tBLASTx, 45.2%
293 identity, e-value $5.51E^{-19}$, supplementary Table 7) with the pestivirus ribonuclease T2 gene.
294 However, considering that this enzyme exists in a wide range of organisms (Luhtala and Parker,
295 2010), the virus-derived origin of this sequence in *Crocidura indochinensis* is not guaranteed.
296 The position of all pestivirus-like hits, including the ribonuclease T2 gene, in the host contig
297 corresponds to their relative organization in the pestivirus genome. No additional features were
298 detected after a tBLASTx search of the whole contig encompassing the two identified EVEs
299 (Supplementary Table S7). In addition, we did not detect the EVE sequences in reads of
300 publicly available experimental genomic and transcriptomic data from Soricidae species.
301

contig accession no.	contig length (nt)	EVE	GenBank accession no.	EVEs fragments	EVE length (nt)	position in host contig		translation frame	Closest BLAST hit	Conserved-domain search
						start	end			
PVKC010097735.1	6104	EVE1	BK014483	No. 1	318	2329	2646	1	Linda virus	Pestivirus envelope glycoprotein E2
		EVE2		No. 2	1053	3205	4257	1	Classical swine fever virus	Peptidase_C74: Pestivirus NS2 peptidase
				No. 3	84	4274	4357	2	Rodent pestivirus	Peptidase_S31: Pestivirus NS3 polyprotein peptidase S31
				No. 4	114	4359	4472	2	Rodent pestivirus	Peptidase_S31: Pestivirus NS3 polyprotein peptidase S31
				No. 5	87	4477	4563	1	Rodent pestivirus	Peptidase_S31: Pestivirus NS3 polyprotein peptidase S31

302 **Table 1. Newly detected EVEs in *Crocidura indochinensis* genome**
303

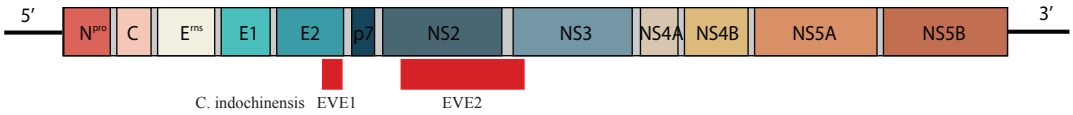

Fig. 1. The positions of the newly detected *Crocidura indochinensis* EVEs are shown relative to an archetypal *Pestivirus* genome (classical swine fever virus, NC_002657). N^{pro}, N-terminal protease; C, nucleocapsid core protein; E^{rns}, envelope glycoprotein E^{rns}; E1, envelope glycoprotein E1; E2, envelope glycoprotein E2; p7, nonstructural protein p7; NS2, nonstructural protein NS2; NS3, nonstructural protein NS3; NS4A, nonstructural protein NS4A; NS4B, nonstructural protein NS4B; NS5A, nonstructural protein NS5A; NS5B, nonstructural protein NS5B.

3.3 – Identification and distribution of pesti-like EVEs in other Soricidae species

To expand our screening and evaluate the distribution of these newly identified EVEs in additional species that are phylogenetically close to *Crocidura indochinensis*, we undertook a PCR-based screening of 65 samples from 29 species of the Soricidae family (Supplementary Table S4). These samples belonged to 7 different genera (*Crocidura*, *Paracrocidura*, *Scutisorex*, *Suncus*, *Sylvisorex*, *Neomys,* and *Sorex*), encompassing two subfamilies, Crocidurinae and Soricinae. Cytb genomic sequences were also generated to confirm the species identification (Supplementary Table S6 & Supplementary Fig. 3).

In total, 58 samples derived from 27 species contained the newly identified pesti-like EVEs, and 48 samples yielded complete or nearly complete EVEs sequences, representing 22 species in the Crocidurinae subfamily and one species (*Neomys anomalus*) in the Soricinae subfamily (Fig. 2). All novel EVEs sequences were highly similar to the *Crocidura indochinensis* EVEs sequences, with a mean identity of 90.38% on amino acid level, and 95.12% on nucleotide level, respectively (Supplementary Table S8). Phylogenetic reconstruction indicated that all Crocidurinae pesti-like EVEs clustered together as a sister-lineage of currently recognized *Pestivirus* species (Fig. 3, Supplementary Fig. 4).

Though collected in different locations, nearly all species tested in the Crocidurinae subfamily harbored the pesti-like EVEs sequences. For some species however, such as *Crocidura* cf. *zimmermanni*, *C. leucodon*, *C. suaveolens* and *Suncus etruscus*, we could not always detect or sequence the EVEs in all samples. The failed detection could be explained by the genomic template being of poor quality due to storage conditions associated with the museum specimens. Interestingly, the pesti-like EVEs were also detected in one *Neomys anomalus* sample, while the remaining Soricinae specimens yielded negative results. The widespread nature of these homolog EVEs in the Crocidurinae species suggests a single endogenization event before their common ancestor about 10.8 million years ago (Dubey et al., 2007). Since we did not manage to sequence the pesti-like EVEs found in *Neomys anomalus*, their phylogenetic relationship to

337 the Crocidurinae pesti-like EVEs is unclear and may not necessarily derive from the same
338 endogenization event. To compare the evolutionary history between the EVEs and the Cytb
339 gene, we constructed a tanglegram based on their respective phylogenies, which illustrates
340 specific phylogenetic discordances that may also be expected between mitochondrial and
341 nuclear genes (Fig 4).
342
343 **3.4 – Selective pressure on pesti-like EVE1 locus**
344 For testing the indication of function of the EVEs region, we further processed selective
345 pressure analysis. We only characterized the selective pressure acting on the EVE1 locus, as
346 the EVE2 locus exhibits widespread stop codons and translation frame shifts. We measured
347 the selective pressure acting on EVE1 by estimating the ratio (ω) of non-synonymous
348 substitution rate (d_N) and synonymous substitutions rate (d_S) in protein-coding sequences using
349 two methods: 1) fixed effects likelihood (FEL) (Kosakovsky Pond and Frost, 2005), and 2)
350 Bayesian renaissance counting (Lemey et al., 2012). It is expected that non-functional region
351 should conform to neutral evolution while functional region would appear to be under purifying
352 selection. Based on the 48 pesti-like EVE1 potentially coding sequences the FEL method
353 indicates an overall neutral evolution with ω = 0.94. Likewise, the Bayesian renaissance
354 counting model yields a ratio at 1.098 [95% credible interval: 0.658, 1.624], reflecting neutral
355 evolution.

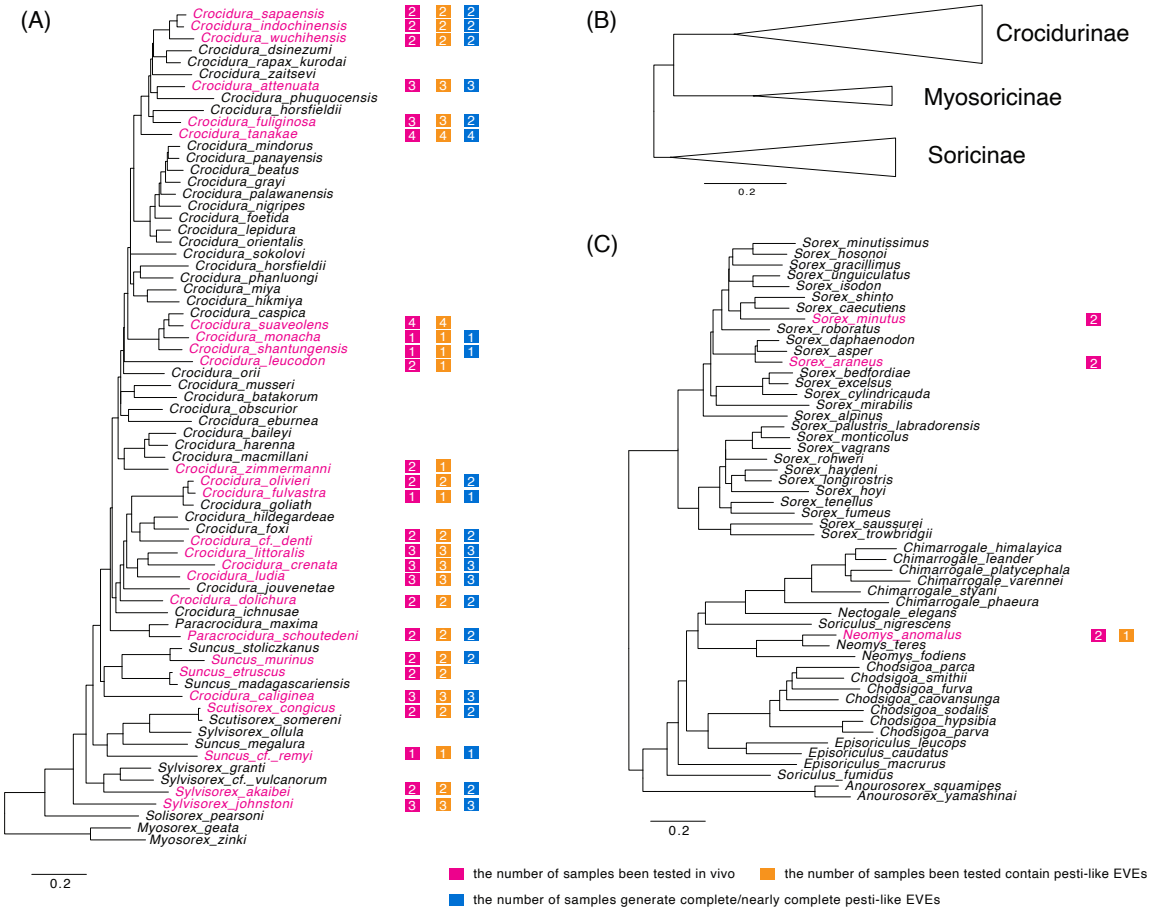

Fig. 2. Maximum likelihood phylogeny of the Soricidae (shrews) family based on the Cytb gene and the available samples' species distribution in this study, highlighted in pink: (A) The phylogeny of Crocidurinae subfamily and sample distribution; (B) Subfamilies relationships within the Soricidae family; (C) The phylogeny of Soricinae subfamily and sample distribution.

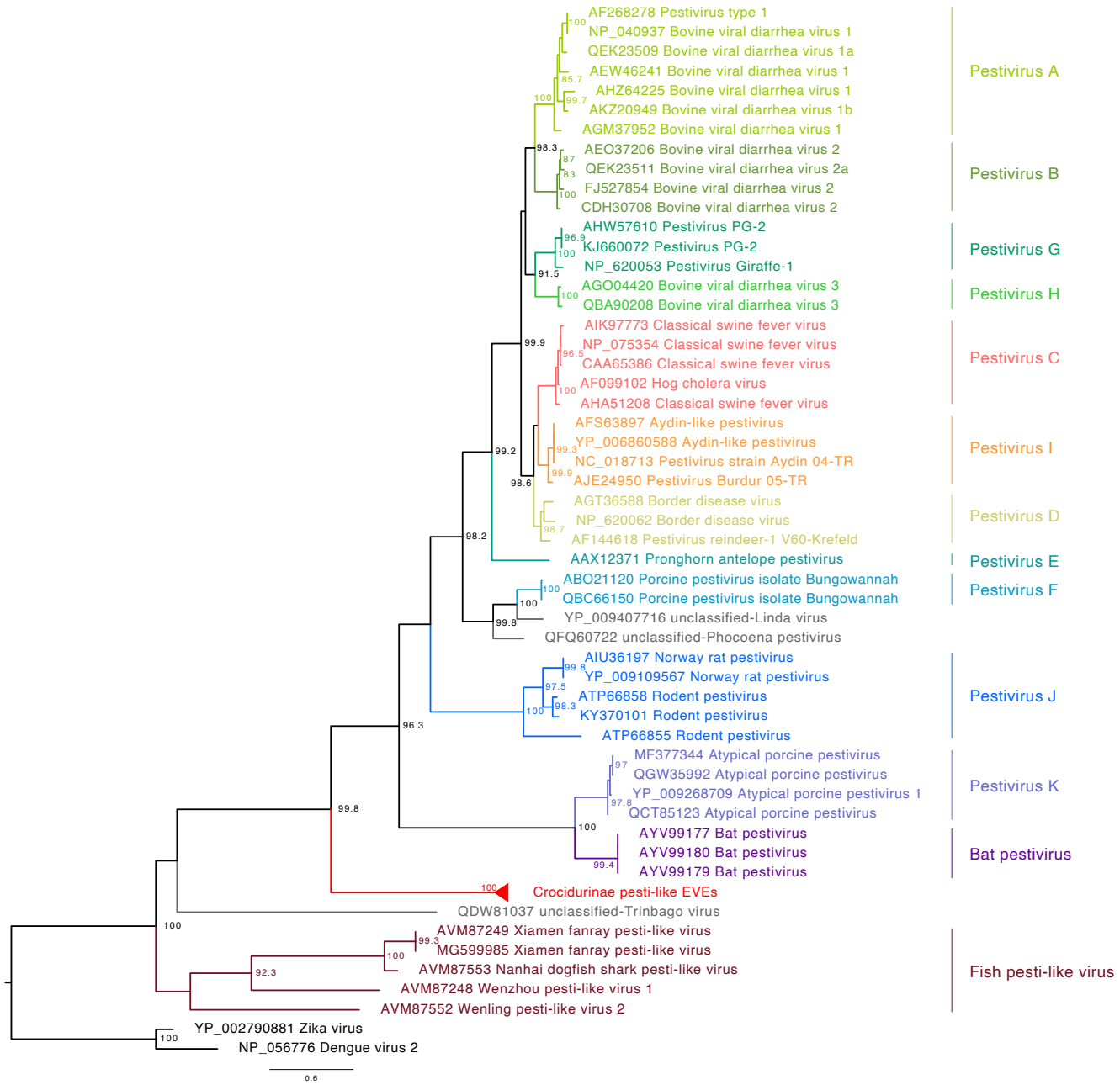

Fig. 3. Phylogenetic relationships of pesti-like EVEs with representative *Pestivirus* species and with Dengue and Zika virus (*Flavivirus*) as outgroup. Clades are colored based on viral species. Node labels indicate Shimodaira-Hasegawa (SH)-like branch support (%, only values > 80% are shown). Scale bars indicate the number of amino acid substitutions per site.

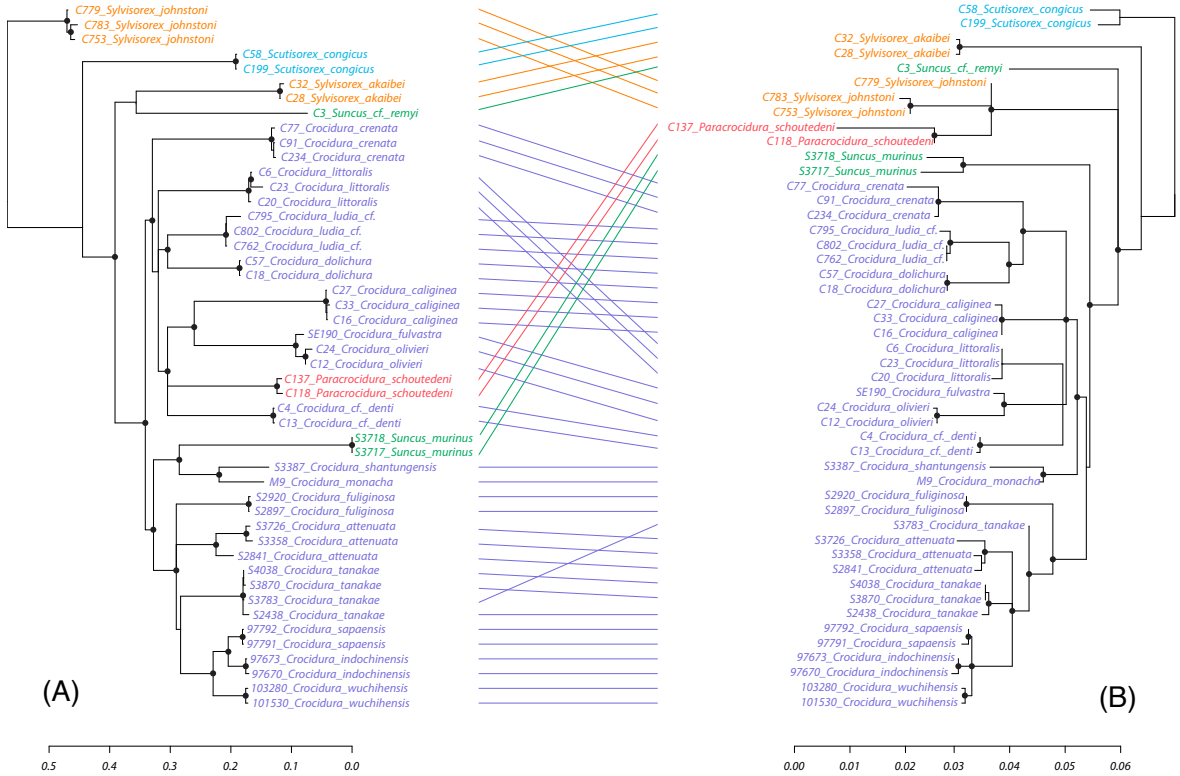

Fig. 4 Tanglegram of the cytochrome b phylogeny (A) and the corresponding EVEs phylogeny (B). The cytochrome b tree was inferred for gene sequences from 22 shrew species, and the EVEs tree (right) was inferred using the newly generated EVEs sequences. The clade nodes with SH-like branch support < 50% were collapsed as polytomies. The black circles indicate the SH-like branch support > 80%. Lines connect corresponding tips in the two phylogenies. Scale bars indicate the number of nucleotide substitutions per site.

4 – Discussion

Non-retroviral endogenous viral elements (EVEs) are rare traces of the ancient evolutionary history of viruses. These genomic fossils offer valuable insights into host range, ancestral genetic diversity and can provide invaluable information for dating viral evolutionary history (Aiewsakun and Katzourakis, 2015a; Feschotte and Gilbert, 2012). In our study we screened a comprehensive set of mammalian genomes to discover such *Flaviviridae*-derived EVEs. We uncovered two *Flaviviridae*-derived EVEs sequences in the genome of the Indochinese shrew and confirmed their presence in a broad range of shrew species belonging to the Crocidurinae subfamily.

The EVEs we identified are related to extant viruses within the *Pestivirus* genus. Viruses belonging to this genus were initially detected in a variety of artiodactylous hosts, such as ruminants and swine, in which they cause subclinical or clinical infections including hemorrhagic syndrome, abortion, acute fatal mucosal disease. Recent metagenomic studies extended the host range towards rodents (Wu et al., 2020, 2018), bats (Wu et al., 2018), fish

386 (Shi et al., 2018), and ticks (Sameroff et al., 2019), but to some extent, the restricted sampling
387 beyond agriculturally important animals limits our understanding of the real host range. Shrews,
388 for example, have been recently identified as host of hepaciviruses, another genus in the
389 *Flaviviridae* family (Guo et al., 2019; Wu et al., 2020), but to date not of pestiviruses. The
390 broad detection of pestivirus-derived EVEs reported in our study strongly supports that
391 ancestors of the Crocidurinae shrew subfamily have been hosts of pestiviruses and suggests
392 that their descendants might still be. Indeed, considering the extremely low probability of a
393 non-retroviral endogenization event to occur in the germline, EVEs are strong indicators of
394 frequent interactions between the original exogenous viruses and their hosts (Aiewsakun and
395 Katzourakis, 2015a; Feschotte and Gilbert, 2012). Although direct detection and
396 characterization of pestiviruses from shrews are still required to formally demonstrate that they
397 are natural hosts of pestiviruses, our study provides indirect support for a wider and more
398 diverse host range of pestiviruses.
399
400 Given the low probability of endogenization events of non-retroviral RNA viruses and the
401 contiguous nature of the two EVEs on the host and viral genome, our results suggest a single
402 endogenization event followed by genetic drift. One or several insertion events separated the
403 original EVE in two fragments, EVE1 and EVE2. EVE1 shows a short but intact open reading
404 frame to be evolving under neutral evolution while EVE2 exhibits multiple stop codons and
405 frame-shifts due to additional insertions. Many studies have identified the important roles that
406 EVEs can play in host antiviral immunity, both in vertebrates and invertebrates (Blair et al.,
407 2020; Ophinni et al., 2019; Skirmuntt et al., 2020). Flavivirus-like EVEs in *Aedes* mosquitoes,
408 for example, can produce P-element-induced wimpy testis (PIWI)-interacting RNAs (piRNAs)
409 which limit the cognate virus replication (Suzuki et al., 2020). It is highly unlikely that the
410 pesti-like sequences we discovered have a function in shrews because of the absence of
411 negative selection and the disruption of the original viral coding region.
412
413 The evolutionary history of the EVEs sequences after integration remains unclear. Interestingly,
414 they do not show any appreciable patterns of phylogenetic consistency with the host Cytb gene
415 sequences. The EVE sequences might not be a reliable genetic marker to discriminate species
416 (Tobe et al., 2010), as highlighted by the low amount of genetic diversity compared to Cytb
417 sequences (Fig 4). Additionally, the discrepancy might be explained by differences in the
418 genetic inheritance of the Cytb gene and EVEs: the Cytb gene is a mitochondrial gene whereas
419 EVEs are integrated in the nuclear genome. This can lead to different observed evolutionary

420 patterns, especially in the case of weak reproductive isolation within species or species
421 complex, allowing hybridization, as has been suggested for some *Crocidura* species (Dubey et
422 al., 2008, 2006; Vogel et al., 2004).
423
424 Dating the ancient evolutionary history of ssRNA viruses such as pestiviruses and *Flaviviridae*
425 in general is challenging. The most commonly used method for inferring viral divergence time
426 is based on the estimation of evolutionary rates derived from sequence data and their collection
427 dates. However, the applicability of this method is often limited by heterogeneous substitution
428 rates though time (Aiewsakun and Katzourakis, 2016) and among viral lineages (Duffy et al.,
429 2008; Sanjuán, 2012). Not accounting for the former leads to recent estimates for the origins
430 of ssRNA viruses that are often in conflict with other phylogenetic evidence (Holmes, 2003).
431 Using suitable molecular clock models, the powerful combination of both tip and node
432 calibrations may help to recover more accurate evolutionary timescales (O'Reilly et al., 2015).
433 Node calibration is however challenging for viruses as no fossil evidence can be found. It thus
434 often relies on known phylogeographic events and other indirect calibrations point, such as
435 ecological events or assumptions of co-divergence as alternative (Bamford et al., 2021;
436 Moureau et al., 2015; Pettersson and Fiz-Palacios, 2014). The discovery of ssRNA virus-
437 related EVEs thus enable a direct estimation for a robust long-term timeline of virus evolution
438 history by co-phyletic analysis of EVE's orthologs in different host (Gilbert and Feschotte,
439 2010).
440
441 The pesti-like EVE sequences characterized in our study are widespread in Crocidurinae
442 species, are monophyletic and exhibit high sequence similarity. Considering the low
443 probability of endogenization events of non-retroviral RNA viruses, this suggests that the pesti-
444 like EVE got integrated before the most-recent common ancestor of the subfamily, which is
445 estimated to be over 10.8 million years ago (Dubey et al., 2007). There are only a handful of
446 molecular dating estimates for pestiviruses and they mostly focus on viral species or clades
447 that are associated with economic losses. Diversification of bovine viral diarrhea virus 1
448 (Pestivirus A) subtypes was estimated to have started about 363 years ago (Weber et al., 2021),
449 and the divergence of HoBi-like pestivirus (Pestivirus H) was dated back to the 16th century
450 (Silveira et al., 2020). Our results provide the first robust ancient time node for pestiviruses
451 based on the estimated EVEs integration date and suggest that pestiviruses were already
452 circulating in mammals more than 10.8 million years ago.

In conclusion, we discovered and characterized the first *Flaviviridae*-related EVEs records from mammalian reference genomes, which derived from pestiviruses. The wide EVEs distribution in shrew Crocidurinae subfamily indicates they are a historical host group of pestiviruses and further suggests a robust ancient origin time of the *Pestivirus* genus. Our results show the key role of EVEs not only in expanding our knowledge about ancient viral-host interactions, but also their importance in reconstructing the viral evolutionary history, which contributes to our understanding of viral evolutionary dynamics from ancient times to the present.

5 – Acknowledgement

The samples from Vietnam were collected during biodiversity surveys carried by the Joint Vietnam-Russian Tropical Research and Technological Centre.

6 – Funding

The research leading to these results has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 725422-ReservoirDOCS). PL acknowledges support by the Research Foundation -- Flanders (`Fonds voor Wetenschappelijk Onderzoek -- Vlaanderen', G066215N, G0D5117N and G0B9317N). YQ LI acknowledges the support of China Scholarship Council. AVA acknowledges the support by Ministry of Science and Higher Education of the Russian Federation (grant no. 075-15-2021-1069). YC LI acknowledges the support of National Natural Science Foundation of China (grant ID. 31672254). F.V.d.P. acknowledges the support of the Ph.D. fellowship from the Research Foundation– Flanders. M.R.W.

476 **References**
477

478 Agapov, E. v., Murray, C.L., Frolov, I., Qu, L., Myers, T.M., Rice, C.M., 2004. Uncleaved
479 NS2-3 Is Required for Production of Infectious Bovine Viral Diarrhea Virus. Journal of
480 Virology 78, 2414–2425. https://doi.org/10.1128/jvi.78.5.2414-2425.2004
481 Aiewsakun, P., Katzourakis, A., 2016. Time-Dependent Rate Phenomenon in Viruses.
482 Journal of Virology 90, 7184–7195. https://doi.org/10.1128/jvi.00593-16
483 Aiewsakun, P., Katzourakis, A., 2015a. Endogenous viruses: Connecting recent and ancient
484 viral evolution. Virology 479–480, 26–37. https://doi.org/10.1016/j.virol.2015.02.011
485 Aiewsakun, P., Katzourakis, A., 2015b. Endogenous viruses: Connecting recent and ancient
486 viral evolution. Virology. https://doi.org/10.1016/j.virol.2015.02.011
487 Aswad, A., Katzourakis, A., 2014. The First Endogenous Herpesvirus, Identified in the
488 Tarsier Genome, and Novel Sequences from Primate Rhadinoviruses and
489 Lymphocryptoviruses. PLoS Genetics 10. https://doi.org/10.1371/journal.pgen.1004332
490 Bannikova, A.A., Abramov, A. v., Borisenko, A. v., Lebedev, V.S., Rozhnov, V. v., 2011.
491 Mitochondrial diversity of the white-toothed shrews (Mammalia, Eulipotyphla,
492 Crocidura) in Vietnam. Zootaxa 1–20. https://doi.org/10.11646/zootaxa.2812.1.1
493 Becher, P., Tautz, N., 2011. RNA recombination in pestiviruses: Cellular RNA sequences in
494 viral genomes highlight the role of host factors for viral persistence and lethal disease.
495 RNA Biology 8, 216–224. https://doi.org/10.4161/rna.8.2.14514
496 Belyi, V.A., Levine, A.J., Skalka, A.M., 2010a. Sequences from Ancestral Single-Stranded
497 DNA Viruses in Vertebrate Genomes: the Parvoviridae and Circoviridae Are More than
498 40 to 50 Million Years Old. Journal of Virology 84, 12458–12462.
499 https://doi.org/10.1128/jvi.01789-10
500 Belyi, V.A., Levine, A.J., Skalka, A.M., 2010b. Unexpected inheritance: Multiple
501 integrations of ancient bornavirus and ebolavirus/marburgvirus sequences in vertebrate
502 Genomes. PLoS Pathogens 6, 1–13. https://doi.org/10.1371/journal.ppat.1001030
503 Blair, C.D., Olson, K.E., Bonizzoni, M., 2020. The widespread occurrence and potential
504 biological roles of endogenous viral elements in insect genomes. Current Issues in
505 Molecular Biology 34, 13–29. https://doi.org/10.21775/CIMB.034.013
506 Bletsa, M., Vrancken, B., Gryseels, S., Boonen, I., Fikatas, A., Li, Y., Laudisoit, A.,
507 Lequime, S., Bryja, J., Makundi, R., Meheretu, Y., Akaibe, B.D., Mbalitini, S.G., van de
508 Perre, F., van Houtte, N., Těšíková, J., Wollants, E., van Ranst, M., Pybus, O.G.,
509 Drexler, J.F., Verheyen, E., Leirs, H., Gouy De Bellocq, J., Lemey, P., 2021. Molecular
510 detection and genomic characterization of diverse hepaciviruses in African rodents.
511 Virus Evolution 7, 1–14. https://doi.org/10.1093/ve/veab036
512 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden,
513 T.L., 2009. BLAST+: Architecture and applications. BMC Bioinformatics 10, 1–9.
514 https://doi.org/10.1186/1471-2105-10-421
515 Colmant, A.M.G., Hobson-Peters, J., Bielefeldt-Ohmann, H., Hurk, A.F. van den, Hall-
516 Mendelin, S., Chow, W.K., Johansen, C.A., Fros, J., Simmonds, P., Watterson, D.,
517 Cazier, C., Etebari, K., Asgari, S., Schulz, B.L., Beebe, N., Vet, L.J., Piyasena, T.B.H.,
518 Nguyen, H.-D., Barnard, R.T., Halla, R.A., 2017. A New Clade of Insect-Specific
519 Flaviviruses from Australian Anopheles Mosquitoes Displays Species-Specific Host
520 Restriction 2, 1–19.
521 Criscuolo, A., Gribaldo, S., 2010. BMGE (Block Mapping and Gathering with Entropy): A
522 new software for selection of phylogenetic informative regions from multiple sequence
523 alignments. BMC Evolutionary Biology 10. https://doi.org/10.1186/1471-2148-10-210
524 Crochu, S., Cook, S., Attoui, H., Charrel, R.N., de Chesse, R., Belhouchet, M., Lemasson,
525 J.J., de Micco, P., de Lamballerie, X., 2004. Sequences of flavivirus-related RNA

viruses persist in DNA form integrated in the genome of Aedes spp. mosquitoes. Journal of General Virology 85, 1971–1980. https://doi.org/10.1099/vir.0.79850-0

de Lamballerie, X., Crochu, S., Billoir, F., Neyts, J., de Micco, P., Holmes, E.C., Gould, E.A., 2002. Genome sequence analysis of Tamana bat virus and its relationship with the genus Flavivirus. Journal of General Virology 83, 2443–2454. https://doi.org/10.1099/0022-1317-83-10-2443

de Perre, F. van, Leirs, H., Cigar, J., Mbalitini, S.G., Itoka, J.C.M., Verheyen, E., 2019. Shrews (Soricidae) of the lowland forests around kisangani (DR Congo). Biodiversity Data Journal 7. https://doi.org/10.3897/BDJ.7.E46948

Dubey, S., Diker, E., Kurtonur, C., Vogel, P., 2008. Secondary contact zones and hybridizations: The case of the lesser white-toothed shrew (Crocidura suaveolens group, Soricidae). Biological Journal of the Linnean Society 95, 557–565. https://doi.org/10.1111/j.1095-8312.2008.01070.x

Dubey, S., Salamin, N., Ohdachi, S.D., Barrière, P., Vogel, P., 2007. Molecular phylogenetics of shrews (Mammalia: Soricidae) reveal timing of transcontinental colonizations. Molecular Phylogenetics and Evolution 44, 126–137. https://doi.org/10.1016/j.ympev.2006.12.002

Dubey, S., Zaitsev, M., Cosson, J.F., Abdukadier, A., Vogel, P., 2006. Pliocene and Pleistocene diversification and multiple refugia in a Eurasian shrew (Crocidura suaveolens group). Molecular Phylogenetics and Evolution 38, 635–647. https://doi.org/10.1016/j.ympev.2005.11.005

Duffy, S., Shackelton, L.A., Holmes, E.C., 2008. Rates of evolutionary change in viruses: Patterns and determinants. Nature Reviews Genetics 9, 267–276. https://doi.org/10.1038/nrg2323

Feschotte, C., Gilbert, C., 2012. Endogenous viruses: Insights into viral evolution and impact on host biology. Nature Reviews Genetics 13, 283–296. https://doi.org/10.1038/nrg3199

Flynn, P.J., Moreau, C.S., 2019. Assessing the diversity of endogenous viruses throughout ant genomes. Frontiers in Microbiology 10. https://doi.org/10.3389/fmicb.2019.01139

G. G.Bamford, C., de Souza, W., Parry, R., J. Gifford, R., 2021. Comparative analysis of genome-encoded viral sequences reveals the evolutionary history of Flaviviridae. bioRxiv 2021.09.19.460981 (preprint).

Galili, T., 2015. dendextend: An R package for visualizing, adjusting and comparing trees of hierarchical clustering. Bioinformatics 31, 3718–3720. https://doi.org/10.1093/bioinformatics/btv428

Gilbert, C., Feschotte, C., 2010. Genomic fossils calibrate the long-term evolution of hepadnaviruses. PLoS Biology 8. https://doi.org/10.1371/journal.pbio.1000495

Guo, H., Cai, C., Wang, B., Zhuo, F., Jiang, R., Wang, N., Li, B., Zhang, W., Zhu, Y., Fan, Y., Chen, Wushen, Chen, Weihong, Yang, X., Shi, Z., 2019. Novel hepacivirus in Asian house shrew, China. Science China Life Sciences 62, 701–704. https://doi.org/10.1007/s11427-018-9435-7

Hartlage, A.S., 2016. The Strange, Expanding World of Animal Hepaciviruses. Annu Rev Virol 176, 53–75. https://doi.org/10.1146/annurev-virology-100114-055104.The

Holmes, E.C., 2011. The evolution of endogenous viral elements. Cell Host and Microbe 10, 368–377. https://doi.org/10.1016/j.chom.2011.09.002

Holmes, E.C., 2003. Molecular Clocks and the Puzzle of RNA Virus Origins. Journal of Virology 77, 3893–3897. https://doi.org/10.1128/jvi.77.7.3893-3897.2003

Horie, M., 2019. Interactions among eukaryotes, retrotransposons and riboviruses: Endogenous riboviral elements in eukaryotic genomes. Genes and Genetic Systems 94, 253–267. https://doi.org/10.1266/ggs.18-00049

575 Horie, M., Honda, T., Suzuki, Y., Kobayashi, Y., Daito, T., Oshida, T., Ikuta, K., Jern, P.,
576 Gojobori, T., Coffin, J.M., Tomonaga, K., 2010. Endogenous non-retroviral RNA virus
577 elements in mammalian genomes. Nature 463, 84–87.
578 https://doi.org/10.1038/nature08695
579 Horie, M., Kobayashi, Y., Suzuki, Y., Tomonaga, K., 2013. Comprehensive analysis of
580 endogenous bornavirus-like elements in eukaryote genomes. Philosophical Transactions
581 of the Royal Society B: Biological Sciences 368. https://doi.org/10.1098/rstb.2012.0499
582 Jenkins, P.D., Abramov, A. v., Bannikova, A.A., Rozhnov, V. v., 2013. Bones and genes:
583 Resolution problems in three Vietnamese species of Crocidura (Mammalia,
584 Soricomorpha, Soricidae) and the description of an additional new species. ZooKeys
585 313, 61–79. https://doi.org/10.3897/zookeys.313.4823
586 Kafetzopoulou, L.E., 2019. Metagenomic sequencing at the epicenter of the Nigeria 2018
587 Lassa fever outbreak. Science 176, 139–148.
588 https://doi.org/10.1016/j.physbeh.2017.03.040
589 Katoh, K., Misawa, K., Kuma, K.I., Miyata, T., 2002. MAFFT: A novel method for rapid
590 multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research
591 30, 3059–3066. https://doi.org/10.1093/nar/gkf436
592 Katzourakis, A., Gifford, R.J., 2010a. Endogenous viral elements in animal genomes. PLoS
593 Genetics 6. https://doi.org/10.1371/journal.pgen.1001191
594 Katzourakis, A., Gifford, R.J., 2010b. Endogenous viral elements in animal genomes. PLoS
595 Genetics 6. https://doi.org/10.1371/journal.pgen.1001191
596 Kawasaki, J., Kojima, S., Mukai, Y., Tomonaga, K., Horie, M., 2021. 100-My history of
597 bornavirus infections hidden in vertebrate genomes. Proceedings of the National
598 Academy of Sciences 118. https://doi.org/10.1073/PNAS.2026235118
599 Kobayashi, Y., Shimazu, T., Murata, K., Itou, T., Suzuki, Y., 2019. An endogenous adeno-
600 associated virus element in elephants. Virus Research 262, 10–14.
601 https://doi.org/10.1016/j.virusres.2018.04.015
602 Kosakovsky Pond, S.L., Frost, S.D.W., 2005. Not so different after all: A comparison of
603 methods for detecting amino acid sites under selection. Molecular Biology and
604 Evolution 22, 1208–1222. https://doi.org/10.1093/molbev/msi105
605 Kosakovsky Pond, S.L., Frost, S.D.W., Muse, S. v., 2005. HyPhy: Hypothesis testing using
606 phylogenies. Bioinformatics 21, 676–679. https://doi.org/10.1093/bioinformatics/bti079
607 Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nature
608 Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923
609 Leinonen, R., Sugawara, H., Shumway, M., 2011. The sequence read archive. Nucleic Acids
610 Research 39, 2010–2012. https://doi.org/10.1093/nar/gkq1019
611 Lemey, P., Minin, V.N., Bielejec, F., Pond, S.L.K., Suchard, M.A., 2012. A counting
612 renaissance: Combining stochastic mapping and empirical Bayes to quickly detect
613 amino acid sites under positive selection. Bioinformatics 28, 3248–3256.
614 https://doi.org/10.1093/bioinformatics/bts580
615 Lequime, S., Lambrechts, L., 2017. Discovery of flavivirus-derived endogenous viral
616 elements in Anopheles mosquito genomes supports the existence of Anopheles -
617 associated insect-specific flaviviruses . Virus Evolution 3, vew035.
618 https://doi.org/10.1093/ve/vew035
619 Li, H., 2018. Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics 34,
620 3094–3100. https://doi.org/10.1093/bioinformatics/bty191
621 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis,
622 G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools.
623 Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

624 Li, N., Li, Q., 2015. Identification and characterization of endogenous viral elements for the
625 three key schistosomes of humans. Pakistan journal of pharmaceutical sciences 28, 375–
626 382.
627 Liu, H., Fu, Y., Xie, J., Cheng, J., Ghabrial, S.A., Li, G., Yi, X., Jiang, D., 2012. Discovery
628 of novel dsRNA viral sequences by in silico cloning and implications for viral diversity,
629 host range and evolution. PLoS ONE 7, 1–10.
630 https://doi.org/10.1371/journal.pone.0042147
631 Liu, S., Coates, B.S., Bonning, B.C., 2020. Endogenous viral elements integrated into the
632 genome of the soybean aphid, Aphis glycines. Insect Biochemistry and Molecular
633 Biology 123, 103405. https://doi.org/10.1016/j.ibmb.2020.103405
634 Luhtala, N., Parker, R., 2010. T2 Family ribonucleases: ancient enzymes with diverse roles.
635 Trends in Biochemical Sciences 35, 253–259. https://doi.org/10.1016/j.tibs.2010.02.002
636 Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey,
637 A.R.N., Potter, S.C., Finn, R.D., Lopez, R., 2019. The EMBL-EBI search and sequence
638 analysis tools APIs in 2019. Nucleic Acids Research 47, W636–W641.
639 https://doi.org/10.1093/nar/gkz268
640 Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer,
641 R.C., He, J., Gwadz, M., Hurwitz, D.I., Lanczycki, C.J., Lu, F., Marchler, G.H., Song,
642 J.S., Thanki, N., Wang, Z., Yamashita, R.A., Zhang, D., Zheng, C., Bryant, S.H., 2015.
643 CDD: NCBI's conserved domain database. Nucleic Acids Research 43, D222–D226.
644 https://doi.org/10.1093/nar/gku1221
645 Mares-Guia, M.A.M.D.M., Horta, M.A., Romano, A., Rodrigues, C.D.S., Mendonça, M.C.L.,
646 dos Santos, C.C., Torres, M.C., Araujo, E.S.M., Fabri, A., de Souza, E.R., Ribeiro,
647 R.O.R., Lucena, F.P., Junior, L.C.A., da Cunha, R. v., Nogueira, R.M.R., Sequeira, P.C.,
648 de Filippis, A.M.B., 2020. Yellow fever epizootics in non-human primates, Southeast
649 and Northeast Brazil (2017 and 2018). Parasites and Vectors 13, 1–8.
650 https://doi.org/10.1186/s13071-020-3966-x
651 Maruyama, S.R., Castro-Jorge, L.A., Ribeiro, J.M.C., Gardinassi, L.G., Garcia, G.R.,
652 Brandão, L.G., Rodrigues, A.R., Okada, M.I., Abrão, E.P., Ferreira, B.R., da Fonseca,
653 B.A.L., de Miranda-Santos, I.K.F., 2014. Characterisation of divergent flavivirus NS3
654 and NS5 protein sequences detected in Rhipicephalus microplus ticks from Brazil.
655 Memorias do Instituto Oswaldo Cruz 109, 38–50. https://doi.org/10.1590/0074-
656 0276130166
657 Moureau, G., Cook, S., Lemey, P., Nougairede, A., Forrester, N.L., Khasnatinov, M.,
658 Charrel, R.N., Firth, A.E., Gould, E.A., de Lamballerie, X., 2015. New insights into
659 flavivirus evolution, taxonomy and biogeographic history, extended by analysis of
660 canonical and alternative coding sequences e0117849. PLoS ONE 10.
661 https://doi.org/10.1371/journal.pone.0117849
662 Nguyen, L.T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A fast and
663 effective stochastic algorithm for estimating maximum-likelihood phylogenies.
664 Molecular Biology and Evolution 32, 268–274. https://doi.org/10.1093/molbev/msu300
665 Öncü, C., Brinkmann, A., Günay, F., Kar, S., Öter, K., Sarıkaya, Y., Nitsche, A., Linton,
666 Y.M., Alten, B., Ergünay, K., 2018. West Nile virus, Anopheles flavivirus, a novel
667 flavivirus as well as Merida-like rhabdovirus Turkey in field-collected mosquitoes from
668 Thrace and Anatolia. Infection, Genetics and Evolution 57, 36–45.
669 https://doi.org/10.1016/j.meegid.2017.11.003
670 Ophinni, Y., Palatini, U., Hayashi, Y., Parrish, N.F., 2019. piRNA-Guided CRISPR-like
671 Immunity in Eukaryotes. Trends in Immunology 40, 998–1010.
672 https://doi.org/10.1016/j.it.2019.09.003

673 O'Reilly, J.E., dos Reis, M., Donoghue, P.C.J., 2015. Dating Tips for Divergence-Time
674 Estimation. Trends in Genetics 31, 637–650. https://doi.org/10.1016/j.tig.2015.08.001
675 Paradis, E., Schliep, K., 2019. Ape 5.0: An environment for modern phylogenetics and
676 evolutionary analyses in R. Bioinformatics 35, 526–528.
677 https://doi.org/10.1093/bioinformatics/bty633
678 Parry, R., Asgari, S., 2019. Discovery of Novel Crustacean and Cephalopod Flaviviruses:
679 Insights into the Evolution and Circulation of Flaviviruses between Marine Invertebrate
680 and Vertebrate Hosts. Journal of Virology 93, 1–20. https://doi.org/10.1128/jvi.00432-
681 19
682 Pettersson, J.H.O., Fiz-Palacios, O., 2014. Dating the origin of the genus Flavivirus in the
683 light of Beringian biogeography. Journal of General Virology 95, 1969–1982.
684 https://doi.org/10.1099/vir.0.065227-0
685 Quinlan, A.R., Hall, I.M., 2010. BEDTools: A flexible suite of utilities for comparing
686 genomic features. Bioinformatics 26, 841–842.
687 https://doi.org/10.1093/bioinformatics/btq033
688 Roiz, D., Vázquez, A., Seco, M.P.S., Tenorio, A., Rizzoli, A., 2009. Detection of novel insect
689 flavivirus sequences integrated in Aedes albopictus (Diptera: Culicidae) in Northern
690 Italy. Virology Journal 6, 1–6. https://doi.org/10.1186/1743-422X-6-93
691 Sameroff, S., Tokarz, R., Charles, R.A., Jain, K., Oleynik, A., Che, X., Georges, K.,
692 Carrington, C. v., Lipkin, W.I., Oura, C., 2019. Viral Diversity of Tick Species
693 Parasitizing Cattle and Dogs in Trinidad and Tobago. Scientific Reports 9, 1–10.
694 https://doi.org/10.1038/s41598-019-46914-1
695 Sanjuán, R., 2012. From molecular genetics to phylodynamics: Evolutionary relevance of
696 mutation rates across viruses. PLoS Pathogens 8, 1–5.
697 https://doi.org/10.1371/journal.ppat.1002685
698 Shi, M., Lin, X.-D., Chen, X., Tian, J.-H., Chen, L.-J., Li, K., Wang, W., Eden, J.-S., Shen,
699 J.-J., Liu, L., Holmes, E.C., Zhang, Y.-Z., 2018. The evolutionary history of vertebrate
700 RNA viruses. Nature 561, E6–E6. https://doi.org/10.1038/s41586-018-0310-0
701 Shi, M., Lin, X.-D., Vasilakis, N., Tian, J.-H., Li, C.-X., Chen, L.-J., Eastwood, G., Diao, X.-
702 N., Chen, M.-H., Chen, X., Qin, X.-C., Widen, S.G., Wood, T.G., Tesh, R.B., Xu, J.,
703 Holmes, E.C., Zhang, Y.-Z., 2016. Divergent Viruses Discovered in Arthropods and
704 Vertebrates Revise the Evolutionary History of the Flaviviridae and Related Viruses.
705 Journal of Virology 90, 659–669. https://doi.org/10.1128/jvi.02036-15
706 Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam,
707 H., Remmert, M., Söding, J., Thompson, J.D., Higgins, D.G., 2011. Fast, scalable
708 generation of high-quality protein multiple sequence alignments using Clustal Omega.
709 Molecular Systems Biology 7. https://doi.org/10.1038/msb.2011.75
710 Silva, E., Marques, S., Osório, H., Carvalheira, J., Thompson, G., 2012. Endogenous
711 Hepatitis C Virus Homolog Fragments in European Rabbit and Hare Genomes Replicate
712 in Cell Culture. PLoS ONE 7. https://doi.org/10.1371/journal.pone.0049820
713 Silva, E., Osório, H., Thompson, G., 2015. Hepatitis C-like viruses are produced in cells from
714 rabbit and hare DNA. Scientific Reports 5, 1–10. https://doi.org/10.1038/srep14535
715 Silveira, S., Cibulski, S.P., Junqueira, D.M., Mósena, A.C.S., Weber, M.N., Mayer, F.Q.,
716 Canal, C.W., 2020. Phylogenetic and evolutionary analysis of HoBi-like pestivirus:
717 Insights into origin and dispersal. Transboundary and Emerging Diseases 67, 1909–
718 1917. https://doi.org/10.1111/tbed.13520
719 Skirmuntt, E.C., Escalera-Zamudio, M., Teeling, E.C., Smith, A., Katzourakis, A., 2020. The
720 Potential Role of Endogenous Viral Elements in the Evolution of Bats as Reservoirs for
721 Zoonotic Viruses. Annual Review of Virology 7, 103–119.
722 https://doi.org/10.1146/annurev-virology-092818-015613

723 Soto, E., Camus, A., Yun, S., Kurobe, T., Leary, J.H., Rosser, T.G., Dill-Okubo, J.A.,
724 Nyaoke, A.C., Adkison, M., Renger, A., Ng, T.F.F., 2020. First Isolation of a Novel
725 Aquatic Flavivirus from Chinook Salmon (Oncorhynchus tshawytscha) and Its In Vivo
726 Replication in a Piscine Animal Model . Journal of Virology 94.
727 https://doi.org/10.1128/jvi.00337-20
728 Strand, T.M., Lundkvist, Å., Olsen, B., Gustafsson, L., 2018. Breeding consequences of
729 flavivirus infection in the collared flycatcher. BMC Evolutionary Biology 18, 1–9.
730 https://doi.org/10.1186/s12862-018-1121-5
731 Suchard, M.A., Lemey, P., Baele, G., Ayres, D.L., Drummond, A.J., Rambaut, A., 2018.
732 Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus
733 Evolution 4, 1–5. https://doi.org/10.1093/ve/vey016
734 Suzuki, Y., Baidaliuk, A., Miesen, P., Frangeul, L., Crist, A.B., Merkling, S.H., Fontaine, A.,
735 Lequime, S., Moltini-Conclois, I., Blanc, H., van Rij, R.P., Lambrechts, L., Saleh, M.C.,
736 2020. Non-retroviral Endogenous Viral Element Limits Cognate Virus Replication in
737 Aedes aegypti Ovaries. Current Biology 30, 3495-3506.e6.
738 https://doi.org/10.1016/j.cub.2020.06.057
739 Tamura, K., Stecher, G., Kumar, S., 2021. MEGA11: Molecular Evolutionary Genetics
740 Analysis Version 11. Molecular Biology and Evolution 38, 3022–3027.
741 https://doi.org/10.1093/molbev/msab120
742 Taylor, D.J., Leach, R.W., Bruenn, J., 2010. Filoviruses are ancient and integrated into
743 mammalian genomes. BMC Evolutionary Biology 10. https://doi.org/10.1186/1471-
744 2148-10-193
745 Tobe, S.S., Kitchener, A.C., Linacre, A.M.T., 2010. Reconstructing mammalian phylogenies:
746 A detailed comparison of the cytochrome b and cytochrome oxidase subunit i
747 mitochondrial genes. PLoS ONE 5. https://doi.org/10.1371/journal.pone.0014156
748 van de Perre, F., Willig, M.R., Presley, S.J., Andemwana, F.B., Beeckman, H., Boeckx, P.,
749 Cooleman, S., de Haan, M., de Kesel, A., Dessein, S., Grootaert, P., Huygens, D.,
750 Janssens, S.B., Kearsley, E., Kabeya, P.M., Leponce, M., van den Broeck, D., Verbeeck,
751 H., Würsten, B., Leirs, H., Verheyen, E., 2018. Reconciling biodiversity and carbon
752 stock conservation in an Afrotropical forest landscape. Science Advances 4.
753 https://doi.org/10.1126/sciadv.aar6603
754 Vogel, P., Maddalena, T., Sarà, M., 2004. Crocidura cossyrensis Contoli, 1989 (Mammalia,
755 Soricidae): Karyotype, biochemical genetics and hybridization experiments. Revue
756 Suisse de Zoologie 111, 925–934. https://doi.org/10.5962/bhl.part.80277
757 Wang, L.G., Lam, T.T.Y., Xu, S., Dai, Z., Zhou, L., Feng, T., Guo, P., Dunn, C.W., Jones,
758 B.R., Bradley, T., Zhu, H., Guan, Y., Jiang, Y., Yu, G., 2020. Treeio: An R Package for
759 Phylogenetic Tree Input and Output with Richly Annotated and Associated Data.
760 Molecular Biology and Evolution 37, 599–603. https://doi.org/10.1093/molbev/msz240
761 Weber, M.N., Wolf, J.M., da Silva, M.S., Mosena, A.C.S., Budaszewski, R.F., Lunge, V.R.,
762 Canal, C.W., 2021. Insights into the origin and diversification of bovine viral diarrhea
763 virus 1 subtypes. Archives of Virology 166, 607–611. https://doi.org/10.1007/s00705-
764 020-04913-y
765 Wu, Z., Han, Y., Liu, B., Li, H., Zhu, G., Latinne, A., Dong, J., Sun, L., Du, J., Zhou, S.,
766 Chen, M., Kritiyakan, A., Jittapalapong, S., Chaisiri, K., Buchy, P., Duong, V., Yang, J.,
767 Jiang, J., Xu, X., Zhou, H., Yang, F., Morand, S., Daszak, P., Jin, Q., 2020. Decoding
768 the RNA viromes of rodent lungs provides new visions into the origin and evolution
769 pattern of rodent-borne diseases in Mainland Southeast Asia.
770 https://doi.org/10.21203/rs.3.rs-17323/v1
771 Wu, Z., Liu, B., Du, J., Zhang, J., Lu, L., Zhu, G., Han, Y., Su, H., Yang, L., Zhang, S., Liu,
772 Q., Jin, Q., 2018. Discovery of diverse rodent and bat pestiviruses with distinct genomic

and phylogenetic characteristics in several Chinese provinces. Frontiers in Microbiology 9, 1–8. https://doi.org/10.3389/fmicb.2018.02562

Yu, G., Smith, D.K., Zhu, H., Guan, Y., Lam, T.T.Y., 2017. Ggtree: an R Package for Visualization and Annotation of Phylogenetic Trees With Their Covariates and Other Associated Data. Methods in Ecology and Evolution 8, 28–36. https://doi.org/10.1111/2041-210X.12628