



HAL
open science

Rabies surveillance in Senegal over fourteen years uncovers first infection of a honey-badger

Martin Faye, Oumar Faye, Nicholas Di Paola, Marie Henriette Dior Ndione,
Moussa Moise Diagne, Cheikh Tidiane Diagne, Ndeye Sakha Bob, Gamou
Fall, Jean-michel Heraud, Amadou Alpha Sall, et al.

► **To cite this version:**

Martin Faye, Oumar Faye, Nicholas Di Paola, Marie Henriette Dior Ndione, Moussa Moise Diagne, et al.. Rabies surveillance in Senegal over fourteen years uncovers first infection of a honey-badger. *Transboundary and emerging diseases*, In press, 10.1111/tbed.14465 . hal-03691736

HAL Id: hal-03691736

<https://hal.science/hal-03691736>

Submitted on 9 Jun 2022

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

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



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License



Rabies surveillance in Senegal over fourteen years uncovers first infection of a honey-badger.

Journal:	<i>Transboundary and Emerging Diseases</i>
Manuscript ID	TBED-OA-829-21
Wiley - Manuscript type:	Original Article
Date Submitted by the Author:	05-Jul-2021
Complete List of Authors:	Faye, Martin; Institut Pasteur de Dakar Faye, Oumar; Institut Pasteur de Dakar Paola, Nicholas Di; US Army Medical Research Institute of Infectious Diseases Ndione, Marie; Institut Pasteur de Dakar Diagne, Moussa; Institut Pasteur de Dakar Diagne, Cheikh Tidiane; Institut Pasteur de Dakar Bob, Ndeye Sakha; Institut Pasteur de Dakar Fall, Gamou; Institut Pasteur de Dakar Heraud, Jean-Michel; Institut Pasteur de Dakar Sall, Amadou; Institut Pasteur de Dakar Faye, Ousmane; Institut Pasteur de Dakar
Subject Area:	Emerging diseases, Virus, Zoonosis/Zoonotics, Wildlife
Keywords:	host-shift, honey-badger, molecular epidemiology, rabies virus, Senegal

SCHOLARONE™
Manuscripts

1
2
3 1 **Title: Rabies surveillance in Senegal over fourteen years uncovers first infection of a**
4
5 2 **honey-badger.**

6
7
8 3 **Authors:** Martin Faye^{1*}, Oumar Faye¹, Nicholas Di Paola², Marie Henriette Dior NDione¹,
9 4 Moussa Moise Diagne¹, Cheikh Tidiane Diagne¹, Ndeye Sakha Bob¹, Gamou Fall¹, Jean-Michel
10 5 Heraud¹, Amadou Alpha Sall¹ and Ousmane Faye¹.

11
12
13
14 6 **Affiliations:**

15
16
17 7 ¹Virology Department, Institut Pasteur de Dakar, 36 Avenue Pasteur, 220 Dakar, Senegal.

18
19
20 8 ²Center for Genome Sciences, United States Army Medical Research Institute of Infectious
21
22 9 Diseases, Fort Detrick, Frederick, Maryland 21702, USA.

23
24
25
26 10 ***Address for correspondence:** Dr Martin FAYE, Virology department, Institut Pasteur de
27 11 Dakar, 36 Avenue Pasteur, 220 Dakar, Senegal; martin.faye@pasteur.sn.

28
29
30
31
32
33 13 **Acknowledgments**

34
35
36 14 This research received no specific grant from any funding agency in the public, commercial, or
37 15 not-for-profit sectors and was only supported by the Institut Pasteur de Dakar. We acknowledge
38 16 colleagues of virology department at Institut Pasteur de Dakar, Senegal for sharing supportive
39 17 information necessary for establishment and accomplishment of this study.

40
41
42
43
44 18 **Ethics Statement**

45
46
47 19 The authors confirm that the ethical policies of the journal, as noted on the journal's author
48 20 guidelines page, have been adhered to. No ethical approval was required.

49
50
51 21 **Conflict of interest statement**

52
53 22 The authors declared no competing interest

54
55
56 23

24 **Summary**

25 Despite the establishment of Rabies surveillance in animals and humans since 2008, there is a
26 lack of data on its circulation, dynamic of transmission and real burden in Senegal. To better
27 understand the molecular epidemiology of rabies virus in Senegal, we investigated the genetic
28 diversity of eighteen new characterized Senegalese rabies virus sequences collected over
29 fourteen years, including a honey-badger-related isolate.

30 Phylogeographic analyses demonstrated that the Senegalese isolates belong to a monophyletic
31 cluster into the Africa-2 clade and supported two RABV introductions in Senegal from West-
32 African neighbour countries, 36-40 years ago. Our study is noteworthy for reporting on the first
33 characterization of an African honey-badger-related rabies virus which hadn't the N-
34 glycosylation site NKT at position 338-G of the glycoprotein.

35 The identified amino acid polymorphisms found in the Senegalese rabies virus sequences are
36 worthy of further investigations. Although a strong multidisciplinary stepwise cooperation is
37 important for the successful elimination of Rabies in dog populations in Senegal by 2030, the
38 establishment of surveillance in wildlife could be necessary to avoid future re-introductions into
39 domestic hosts.

40 **Keywords:** host-shift, honey-badger, molecular epidemiology, rabies virus, Senegal.

42 **Introduction**

43 Rabies is a neglected, acute as well as progressive neurological disease that is invariably fatal,
44 once symptoms onset, without prompt post-exposure prophylaxis (PEP) (1). Rabies is one of the
45 most important zoonotic diseases and occurs in over 150 countries and territories around the
46 world, causing between 50,000 and 70,000 human deaths annually (2). Most human rabies
47 deaths were associated with bites or scratches from an infected dog, mainly in developing
48 countries in Africa and Asia, where Rabies remains a major public health concern (1, 3).

49 Rabies is caused by rabies virus (RABV), which is a neurotropic and enveloped virus belonging
50 to the *Lyssavirus* genus (order *Mononegavirales*, family *Rhabdoviridae*). The RABV's genome

1
2
3 51 is a negative single-stranded RNA of approximately 12 kilobases (kb) in size, which encodes for
4 52 five structural proteins (N, P, M, G, and L). The nucleoprotein (N), the phosphoprotein (P) and
5 53 the RNA-dependent RNA polymerase form the ribonucleoprotein (RNP), which plays a crucial
6 54 role in RABV transcription and replication (4). The N and L proteins harbor the most conserved
7 55 regions among the RABV genome (4). Known as the key protein involved in virus assembly,
8 56 budding and virions release, the matrix protein (M) is also as a regulator of viral transcription
9 57 and replication (5). The glycoprotein (G) is the only surface protein and plays a major role
10 58 during the initial steps of the infectious cycle. It is also essential for induction of host's innate
11 59 and adaptive immunity (6, 7).

12 60 RABV can be transmitted by a wide range of animal hosts including many species of wild
13 61 carnivora and chiroptera (8).

14 62 Substantial advances in understanding RABV biology (8), pre- and post-exposure treatment (6,
15 63 7), and massive immunization campaigns conducted to the success of eradication and control
16 64 programs in Europe (9) and the Americas (10). Nevertheless, the objective of no Rabies death by
17 65 2030 is difficult to achieve in developing countries despite some RABV-targeted control efforts
18 66 (1). Although human and dog-mediated Rabies eradication is an achievable objective in
19 67 developing countries through a strong One health cooperation, the eradication of Rabies in the
20 68 wild seems more difficult considering that rabies viruses continue to spread establish in
21 69 geographical areas and, new host species such as various species belonging to the genera
22 70 *Melogale*, *Meles*, and *Mellivora* in the weasel family *Mustelidae* (11).

23 71 Sequenced RABV genomes demonstrate expansive intra-species genetic diversification with
24 72 seven major phylogenetic groups circulating throughout the world (). Three of these groups are
25 73 circulating in Africa including the Africa-2, Africa-3 and Cosmopolitan clades (12-14). Previous
26 74 phylogenetic studies based on the N gene sequences have shown that Africa-2 clade circulates in
27 75 Central and West African countries including in Senegal (13). Prior to this work, only two
28 76 complete viral genomes from Senegal were available on GenBank (KX148238-39) and were
29 77 generated from strains isolated in 1991 and 1992, respectively (12).

1
2
3 78 Here, we investigated the phylodynamic of the RABV circulating in Senegal and West Africa,
4
5 79 using complete genomes from eighteen isolates collected over a fourteen-year period to gain a
6
7 80 better understanding of virus spread patterns and timing.
8

9 81 **Materials and Methods**

10 82 *Sample Collection*

11
12
13
14
15 83 All 18 virus isolates analyzed in this study were derived from experimentally infected suckling
16
17 84 mice brain tissues preserved in the collection of the national reference center for RABV at the
18
19 85 Institut Pasteur de Dakar, Senegal (NRC-Rabies IPD) (**Table 1**). One of these strains was
20
21 86 isolated in 2011 from a honey-badger or ratel (*Mellivora capensis*) at Dahra and this is the first
22
23 87 time in Senegal. Dahra is a city of up to 30,000 people, located in the Linguère department,
24
25 88 Louga region in Senegal (15°21'N; 15°36'W), at approximately 264 km from the capital Dakar.
26
27 89 The main activities in Dahra are agriculture and animal breeding.

28
29 90 RABV infection was confirmed by real-time reverse-transcriptase quantitative polymerase chain
30
31 91 reaction (RT-qPCR) and direct fluorescent antibody test (DFAT) as previously described (3, 15)
32
33 92 . Extraction of viral RNA from 140 microliter (μL) of supernatants from suckling mice brain
34
35 93 tissues, was performed with the QIAamp viral RNA mini kit (Qiagen, Heiden, Germany)
36
37 94 according to manufacturer's instructions and eluted in a final volume of 60 μL . Extracted RNA
38
39 95 was stored at $-80\text{ }^{\circ}\text{C}$ prior to downstream applications.

40 96 *Complete genome sequencing*

41
42 97 Reverse-transcription was performed using the Avian Myeloblastosis Virus (AMV) kit
43
44 98 (Promega, Madison, WI, USA) with random primers, following manufacturer's instructions. The
45
46 99 complete polyprotein was obtained using overlapping polymerase chain reactions (PCR) with
47
48 100 each primer set and the GoTaq® DNA polymerase kit (Promega, Madison, WI, USA) in a final
49
50 101 volume of 50 μL , according to manufacturer's instructions. Primers details are indicated in
51
52 102 **Appendix Table 1**. Briefly, five μL (around 10 μg) of cDNA was added to 45 μL of a RT-PCR
53
54 103 mix and PCR was carried out as previously described (16). Subsequently, five μL of each PCR
55
56 104 product was analyzed by gel electrophoresis on 1% agarose gels stained with ethidium bromide
57
58 105 using a DNA molecular weight marker (HyperLadder™ 1 kb, Bioline, Taunton, MA, USA) to

1
2
3 106 check the size of amplified fragments. The DNA amplicons were purified (QIAquick Gel
4 107 Extraction Kit, Qiagen, Heiden, Germany) and sequenced from both ends for each positive
5 108 sample using Sanger (Beckmann Coulter, High Wycombe, UK). Sequencing of the 3' leader and
6 109 5' trailer of the viral genome was performed using a 5' RACE kit (Invitrogen, Carlsbad, CA,
7 110 USA) and a 3' RACE kit (Roche, Basel, Switzerland) following the manufacturer's protocols.

111 ***Sequence Analysis***

112 Overlapping genomic sequences were assembled using the Unipro UGENE software
113 (<http://ugene.net/download.html>) (17) and obtained full-genomes were aligned using Muscle
114 algorithm (<http://www.drive5.com/muscle/>) (18) within Unipro UGENE software. Based on these
115 alignments, the genetic properties such as pairwise nucleotide and amino acid distances at genes
116 level were analyzed among of the new characterized Senegalese sequences and between them
117 and the previously available complete Africa-2 clade genomes from Nigeria (GenBank
118 accession: KC196743) and Central African Republic (GenBank accession: KF977826).

119 In addition, amino acid substitutions were also assessed on the most divergent sequences from
120 the previously available complete RABV genomes from Senegal (KX148238-39) and compared
121 to specific mutations previously described for mongoose-related RABV (Africa-3 clade) and
122 ferret-badger-related RABV (SEA5 sub-clade and the SEA2b lineage) (12).

123 As it represents the major contributor to RABV pathogenicity (19), the glycoprotein (G)
124 sequences from the new characterized isolates were screened for conservation of amino acid
125 residues of virulence previously described (20-22). A motive was considered as non-conserved
126 when it presents more than two non-conservative mutations in more than two RABV isolates. In
127 addition, polymorphisms on antigenic sites of 10 vaccine strain G protein sequences were
128 assessed on the new characterized Senegalese rabies virus sequences.

129 ***Recombination and positive selection detection***

130 The presence of recombination events was assessed in 37 RABV complete genome sequences
131 from Western and Central Africa using seven methods (RDP, GENECONV, MaxChi, BootScan,
132 Chimaera, SiScan, and 3Seq) implemented in the Recombination Detection Program (RDP4.97)
133 (<http://web.cbio.uct.ac.za/~darren/rdp.html>) (23) and the Genetic Algorithms for Recombination

1
2
3 134 Detection (GARD) method implemented in Datamonkey web server (<http://datamonkey.org>)
4
5 135 (24). The settings were kept at their default values. In addition, episodes of positive diversifying
6
7 136 selection were analyzed applying four different methods (SLAC, MEME, FUBAR, aBSREL)
8
9 137 implemented in the HyPhy package from Datamonkey web server (24). An episode of positive
10
11 138 selection was considered if it was detected by at least two different methods.

12 13 139 ***Molecular evolution and phylodynamics***

14
15 140 Additional RABV sequences were downloaded from GenBank; long stretches of ambiguous
16
17 141 bases or sequences derived from experimental infections or constructs were removed. Our initial
18
19 142 dataset included 19 sequences from Western and Central Africa (**Appendix Table 2**).

20 143 Alignments were done in Geneious, version 11.1.4, using MAFFT, version 7.388. For each
21
22 144 RABV alignment, maximum-likelihood phylogenies were generated using PhyML version 3.3
23
24 145 with the GTR+ Γ model with 1000 bootstrap replicates. The maximum-likelihood phylogenies
25
26 146 were used for estimates of root-to-tip distances, regression slopes, and correlations using
27
28 147 TempEst with the best-fitting root option (25). Tip divergence data were exported and mapped
29
30 148 with linear regression (95% prediction interval) in Rstudio, version 3.2.3, with the R *Stats* and
31
32 149 *ggplot2* packages (26).

33 150 Bayesian analysis using BEAST (versions 1.10.0) was used to generate maximum-clade
34
35 151 credibility (MCC) trees. Phylogenetic relationships and evolutionary rates were modeled
36
37 152 according to the Hasegawa-Kishino-Yano 85 (HKY85) model with gamma-distributed rate
38
39 153 variation and invariant rate distribution among sites (HKY85 + Γ_4 + I) (27). Analyses shared a
40
41 154 strict molecular clock, a nonparametric coalescent “SkyGrid” model (number of grid points + 1 =
42
43 155 50) as a prior density across trees, and an approximate continuous time Markov chain rate
44
45 156 reference prior (28, 29). Tracer v.1.6 was used to ensure run convergence (effective sample size
46
47 157 >200) (30). Each chain consisted of up to 1.0×10^8 Markov chain Monte Carlo steps (1.0×10^7
48
49 158 were discarded as burn-in). Parameters and trees were sampled every 100,000 generations.
50
51 159 TreeAnnotator1.8.4 was used to calculate a MCC tree using a posterior probability limit of 0.7,
52
53 160 which removes summary statistics from nodes with low support (31).

54
55 161 To infer the historical movement of RABV across Western Africa, we implemented an
56
57 162 asymmetric continuous-time Markov chain (CTMC) approach using country as a discrete trait.

1
2
3 163 Bayesian stochastic search variable selection (BSSVS) and SpreadD3
4
5 164 (https://rega.kuleuwn.be/cev/ecv/software/Spread3_tutorial) were used to identify strongly
6
7 165 supported migrations (i.e., any with a Bayes factor ≥ 3) (**Appendix Table 3**).
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review Only

166 **Results**

167 ***Genomic characterization***

168 From 2001 to 2015, 18 RABV complete genome sequences were generated from four humans,
169 13 canines, and one honey-badger (*Mellivora capensis*) samples (**Table 1**). As the previously
170 available complete sequences from Senegal (12), the new characterized RABV isolates showed a
171 genome length of 11,923 nt with similar size at protein and non-coding region levels.

172 ***Genetic distances***

173 Pairwise nucleotide and amino acid distances of coding sequences were evaluated between
174 isolates characterized in this study and in comparison, with previously available Senegalese
175 RABV sequences and non-Senegalese sequences. Nucleotide percent (%) dissimilarity of RABV
176 sequences from Senegal comparatively ranged from 0.2% to 1.7% while a nucleotide distance of
177 4% was found between the new characterized sequences from Senegal and previously described
178 isolates from Nigeria (DRV-NG11, GenBank accession: KC196743) and Central African
179 Republic (CAR_11_001h, GenBank accession: KF977826) belonging to the Africa-2 clade (32,
180 33). However, the assessment of amino acid distances at proteins level between the Senegalese
181 RABV sequences showed a dissimilarity ranging between 0 and 2% for the N, P, M and L
182 proteins while the G protein exhibited more variability with an amino acid distance ranging from
183 0% to 7%. The highest amino acid distances in the G protein were observed between the isolate
184 SA267115 (Dakar, SN 2014) from dog and the isolates SA212203 (Dahra, SN 2011) from
185 honey-badger (7%) and SA252888 (Dakar, SN 2013) isolated from dog (7%). These 3 isolates
186 presented also the highest pairwise amino-acid distances from the other Senegalese sequences
187 with ranges between 3-4%, 3-6%, and 5-6% for SA267115, SA212203 and SA252888,
188 respectively. The amino-acid distances between the G protein sequences of the other Senegalese
189 RABV isolates analyzed in this study ranged from 1% to 2%.

190 ***Amino acid substitutions on proteins of the most divergent Senegalese RABV sequences***

191 The most divergent Senegalese RABV sequences from two dog-related RABV isolates
192 (SA267115 and SA252888) and the honey-badger isolate (SA212203) were screened for
193 mutations at genes level in comparison to the other Senegalese RABV sequences. No mutation

1
2
3 194 was found in sequences of these three RABV isolates for the N, P and M proteins. However, a
4
5 195 total of 3 and 9 substitutions were identified for the dog-related isolate SA267115, in the L
6
7 196 protein and the G protein, respectively. The second dog-related RABV isolate SA252888
8
9 197 exhibited no mutation in the polymerase L while a total of 8 amino acid substitutions was found
10
11 198 in its G protein sequence. The RABV sequence from honey-badger showed 8 and 7 amino acid
12
13 199 mutations for the polymerase L and the G protein, respectively. Previously described specific
14
15 200 mongoose-related and ferret-badger-related RABV substitutions (12) were assessed among the
16
17 201 mutations identified in the Glycoprotein and the polymerase of these three Senegalese isolates.
18
19 202 The dog-related RABV isolate SA252888 have showed a mutation (Pro-386G-Leu) quietly
20
21 203 similar to the Pro-386G-Ser mutation previously described as specific to mongoose-related
22
23 204 RABV (Africa-3 clade) (**Table 2**).

23 205 *Amino acid polymorphisms in the G protein*

24
25 206 The presence of 13 highly conserved amino acid residues of virulence was assessed across the G
26
27 207 protein of the new Senegalese RABV sequences using multiple alignments. The identified
28
29 208 conservative mutations were highlighted in black while the non-conservative mutations were
30
31 209 colored in black and underlined (**Table 3**). A total of 10 amino acid residues of virulence among
32
33 210 13 was conserved in the G protein of Senegalese RABV isolates with sometimes non-
34
35 211 conservatives amino acid changes. However, non-conservative mutations Arg333Ala and
36
37 212 Asn194Tyr were identified in all Senegalese RABV sequences. Highly conserved in the
38
39 213 *Lyssavirus* genus, the N-glycosylation site N319 was identified at amino acid position 338 in G
40
41 214 protein of the Senegalese RABV isolates and the previously described motif of glycosylation
42
43 215 NKT (26) showed polymorphisms in some of the new characterized Senegalese sequences.
44
45 216 Indeed, the isolate SA194858 (Dakar, SN 2008) from dog exhibited a mutation Phe340Ala at the
46
47 217 glycosylation site (NKA) while the isolate SA252888 from dog presented two non-conservative
48
49 218 mutations (Lys339Glu and Phe340Ala) at the same site (NEA). In addition, the isolate
50
51 219 SA212203 from honey-badger did not present this N-glycosylation site and exhibited an amino-
52
53 220 acid mutation Asn338Asp (DKT) (**Table 3**).
54
55 221 G protein antigenic sites (22) were relatively conserved with a few exceptions. Isolates
56
57 222 SA194858, SA212203, and SA252888 presented with multiple changes in antigenic site III
58
59 223 (residues 330–338). Comparison with 10 vaccine strain G protein sequences also presents with
60

224 disagreements (**Figure 1**). G protein percent identity between SA252888 and vaccine strains
225 ranged from 84.5% to 87.6%. On average, percent identity of Senegalese G proteins was ~91.0%
226 when compared to the selected vaccine strains.

227 ***Recombination and episodic selection***

228 There was no evidence for recombination detected in any RABV sequences used in this study.
229 Several sites under strong negative selection were found with the SLAC (P value ≤ 0.1) and
230 FUBAR (posterior probability [PP] ≥ 0.9) models and most of them were located in major
231 proteins such as the L, N and G protein, respectively. However, significant episodic positive
232 selection was obtained for all the proteins, except for the polymerase L (**Table 4**). All positively
233 selected sites estimated by the FUBAR model (PP ≥ 0.9) were also identified by the MEME
234 method (P value < 0.1) and the majority of such sites were located in the G protein. The N, P and
235 M protein of the West African RABV sequences exhibited each, 1 amino acid site (N270, P131
236 and M60 respectively) under positive selection while the G protein showed a total of 6 amino
237 acid sites (G144, G236, G318, G321, G347, G458) (P value ≤ 0.1 and PP ≥ 0.9). Two among
238 these sites harbored mutations from the honey-badger-related RABV isolate (Ser-321-Asp and
239 His-347-Trp). Branch-site analysis showed also 1 branch evaluating under positive selection (P
240 value < 0.05) in the Glycoprotein and this episodic selection event was identified in the sequence
241 of the honey-badger-related isolate SA212203. The eight sites detected under positive selection
242 in the polymerase protein involved a previously characterized strain from Guinea (18) and were
243 identified only with the MEME method; then considered as non-evident (**Table 4**).

244 ***Phylogenetics of RABV in Senegal and Western Africa***

245 A maximum-likelihood phylogenetic tree (**Figure 2**) and Bayesian-inferred MCC tree (**Figure 3**)
246 of Senegalese RABV sequences formed a strongly supported monophyletic clade. Additionally,
247 Senegalese RABV sequences share a most recent common ancestor with sequences isolated from
248 Cote d'Ivoire, Burkina Faso, and Mauritania. Due to the weak temporal signal in only
249 Senegalese RABV genomes ($r^2 = 0.20$), phylogeographic predictions included additional African
250 RABV isolates. Altogether, related African RABV sequences showed a stronger temporal signal
251 ($r^2 = 0.64$, **Figure 2A**).

1
2
3 252 Time to the most recent common ancestor (tMRCA) for our RABV dataset was estimated to be
4
5 253 in the early 19th century (95% HPD: 1803-1848, **Figure 3**). Based on currently available
6
7 254 sequences, the tMRCA for the emergence of RABV in Senegal was in the 1980s (95% HPD:
8
9 255 1981-1986). Patterns of virus evolution were used to infer a molecular epidemiology that
10
11 256 supports two RABV introductions in Senegal from West-African neighbour countries. Moderate
12
13 257 evidence points to two migration events: one from Cote d'Ivoire (Bayes factor: 4.6, **Appendix**
14
15 258 **Table 3**) and the other from Burkina Faso (Bayes factor: 3.1, **Appendix Table 3**). Evolutionary
16
17 259 rates of included RABV sequences had mean substitution rate of 2.8×10^{-4} (95% HPD: $2.5-3.2 \times$
18
19 260 10^{-4}).

20 261 **Discussion**

21
22 262 The burden of RABV in Africa continues to hold steady around 20,000 deaths a year (3), with
23
24 263 limited information on zoonotic circulation in West African countries (34). Despite the
25
26 264 establishment of a national surveillance program of Rabies infection in humans and animals
27
28 265 since 2008, by the Institut Pasteur de Dakar and the Senegalese Ministry of Health and Social
29
30 266 Actions, only few complete genomic data were available from Senegal prior to this study (12).
31
32 267 Here, we sequenced 18 RABV complete genome sequences isolated from humans, dogs and a
33
34 268 honey-badger over a 14 year-period. Notably, this is the first documented whole genome
35
36 269 sequence of honey-badger-related RABV in West Africa, which could have implications for
37
38 270 future spillover events in Senegal. We assessed the genetic diversity of RABV from Senegal by
39
40 271 evaluating amino-acid distances at protein level and identifying mismatches in motives of
41
42 272 virulence and antigenic sites located in the glycoprotein. In addition, we estimated the episodic
43
44 273 selection pressures at proteins level and the phylodynamic analysis of RABV circulation in West
45
46 274 Africa.

47 275 Overall, the new Senegalese RABV sequences are closer to previously available genomes from
48
49 276 Senegal (12). The low pairwise nucleotide distances (< 2%) observed between the Senegalese
50
51 277 RABV sequences has resulted in a more apparent diversity at amino acid level (< 3%). The low
52
53 278 amino acid sequence diversity observed in major proteins of the Senegalese RABV isolates such
54
55 279 as the Nucleoprotein and the polymerase L, confirmed the description of these two regions as the
56
57 280 most conserved among Lyssaviruses (4). However, three isolates showed the highest amino-acid
58
59 281 distances between them and from the other Senegalese sequences of the G protein. The G

1
2
3 282 protein-specific diversity exhibited in the Senegalese sequences could have serve as a factor for
4
5 283 virus spillover in a new host such as honey-badger (13). In addition, the higher numbers of
6
7 284 amino acid mutations from the three most divergent isolates, were found in the Glycoprotein.
8
9 285 Interestingly, the identification of a mutation closed to a mongoose-specific RABV's substitution
10
11 286 from a dog-related RABV sequence in this study (Pro-386G-Leu), shows that the Pro-386G-Ser
12
13 287 could be not only present in mongoose-related Africa-3 RABV (35), but also shared between
14
15 288 mongoose and dog-related isolates. Future studies could be conducted to investigate the
16
17 289 possibility of RABV circulation in a wide variety of natural and incidental animal host species in
18
19 290 Senegal. Additionally, how species or isolate specific changes play a role in PEP efficiency
20
21 291 should also be considered. Furthermore, the substitutions observed in the polymerase protein
22
23 292 could have participated to the RABV shift to new host species in Senegal such as honey-badger.
24
25 293 Despite several isolations of RABV from honey-badger in Southern Africa (36), no associated
26
27 294 sequence was available. Our study is then noteworthy not only for generating the first whole
28
29 295 genome sequence of a honey-badger-related RABV, but also for identifying the specific amino
30
31 296 acid substitutions characterizing RABV isolated from the African honey-badger.

30 297 The G protein plays a crucial role in the biology and pathogenesis of neurotropic RABV
31
32 298 infection (6, 7). The amino acid mutation Arg-333-Ala have been shown to abolish virulence
33
34 299 while the substitution Asn-194-Tyr is implicated in the ability of RABV to infect neurons (19,
35
36 300 37). These two substitutions in the Glycoprotein were highly associated to RABV
37
38 301 neuropathogenesis in furious and paralytic forms of Rabies (38). Previously described as one of
39
40 302 the two most conserved N-glycosylation sites among sequences of the street RABV strains and
41
42 303 the most stable site in the expression of the glycoprotein (21), the glycosylation site at position
43
44 304 319 (Asn319) was located at amino acid position 338 in the Senegalese sequences. Although the
45
46 305 amino acid substitution at position 338 of the G protein plays no major role in viral
47
48 306 pathogenicity, the identification of a glycosylation site at this position could be a new insight in
49
50 307 the G-protein specific evolution of RABV in Senegal (39, 40). More noteworthy, our findings
51
52 308 highlighted also the absence of this N-glycosylation motive (NKT) in sequence of the RABV
53
54 309 isolate from honey-badger. Thus, further studies using the recombinant technology, could be
55
56 310 needed to investigate the effect of this motive in the pathogenicity of the Senegalese isolates
57
58 311 (42).

1
2
3 312 Although low levels of recombination have not been reported for RABV (41), natural selection is
4 313 believed to be one of the drivers of their evolution (12). RABV is subject to strong purifying
5 314 selection and the lowest nucleotide substitution rates were described in the following order: N, L,
6 315 G, M and P (12). As expected, our analysis also showed that the L and N proteins often
7 316 described as more conserved for RABV, exhibited the strongest purifying selection. These two
8 317 proteins are functionally important in RABV pathogenicity (4). In addition, the absence of
9 318 positive selection in the polymerase L suggests frequent purging of deleterious polymorphisms in
10 319 this region during RABV transmission, since neither of the motives identified in the polymerase
11 320 was found to be subject to positive selection using the methods employed in our study. Beside to
12 321 the other proteins, there is substantial evidence that episodic selection events in the G protein
13 322 sequences are a result of the evolution of RABV in West Africa. The amino acid substitution
14 323 318G (Phe) has been previously associated to furious and paralytic Rabies and involved in p75
15 324 neurotrophin receptor binding, viral maturation and transport into the cell (36). Although the
16 325 other sites subjected to positive selection in the G protein were not documented in RABV
17 326 pathogenicity, almost all functional sites directly related to host tropism and infectivity are
18 327 located from positions 181 to 431 in the of the G protein derived from the carnivore RABV (42).
19 328 The presence of specific mutations under positive diversifying selection in the sequence from the
20 329 honey-badger-related RABV exhibited also the nature of selection pressures associated with host
21 330 switching which could be considered as a pattern driving RABV evolution in West Africa,
22 331 particularly in Senegal (43).

23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39 332 Phylogenetic analysis shows that the Senegalese RABV sequences are grouped into a
40 333 monophyletic cluster (Africa-2 clade) and are not region (intra-Senegal) specific. The ML tree
41 334 topology in this paper is in agreement with previously published data including sequences from
42 335 Senegal (13). However, the spillover of RABV into a novel ecological niche (host shift to the
43 336 honey-badger) could lead to the rapid spreading and future diversification of the virus in
44 337 Senegal. Spillover infections are important public health events that establish new reservoirs for
45 338 human exposures and can affect conservation of threatened or endangered wildlife species when
46 339 they occur into those species (44). Although Combinations of positively selected changes have
47 340 been shown to be associated to genetic polymorphism during each RABV host shift (45).
48 341 Therefore, it is becoming increasingly clear that the adaptive changes necessary for RABV host
49 342 shift is determined in part by viral factors that differ between strains, by the interaction between
50
51
52
53
54
55
56
57
58
59
60

1
2
3 343 the viral factors and the identity of the donor and recipient host species involved (46). Data from
4
5 344 the Senegalese RABV have been previously included in genome-wide phylogenetic analyses
6
7 345 (12). Nevertheless, our results represented the most comprehensive molecular analysis of RABV
8
9 346 from Senegal, as it includes more whole genome sequences than previously documented. The
10
11 347 generated complete sequences could be useful not only for large-scale analysis of RABV
12
13 348 dynamics in Africa, but also in update of the existing pan-Lyssavirus or RABV-specific
14
15 349 diagnostics.

16 350 Despite the host switching experienced since 2011 by the Africa-2 clade circulating in West
17
18 351 Africa through spillover to honey-badger in Senegal, it showed a mean nucleotide substitution
19
20 352 rate (2.8×10^{-4} (95% HPD: $2.5\text{--}3.2 \times 10^{-4}$)) which is in agreement with findings reported in
21
22 353 previous studies of RABV evolution (12, 47). All the RABV sequences generated in this study
23
24 354 formed a homogenous cluster in the MCC tree (nucleotide dissimilarity ranging from 0.2% to
25
26 355 1.7%), suggesting a close-related population of RABV in Senegal as previously described in a
27
28 356 detailed analysis of the phylogeographical structure of the Africa-2 clade between Central to
29
30 357 Western African regions (13). Our data suggest two recent RABV introductions into Senegal
31
32 358 during the 1980s (95% HPD: 1981-1986) from Cote d'Ivoire and Burkina Faso. Instead of a
33
34 359 transboundary animal transmission of Rabies from these countries (13, 48), the phylogenetic
35
36 360 evidence gathered in this study points more in the direction of long-distance transmission of
37
38 361 Rabies facilitated by human-mediated animal movements (48, 49). However, further
39
40 362 phylogenetic studies including more complete data sequences from these two latest countries
41
42 363 could be necessary to provide more consistency to these results as there is only 3 and 2 complete
43
44 364 genome RABV sequences available from Cote d'Ivoire and Burkina Faso, respectively (12).

45 365 In conclusion, regarding the key genetic elements from honey-badger-related RABV, the host
46
47 366 shift has resulted in a slightly increased genetic diversity and slow evolution of the Africa-2
48
49 367 clade RABV circulating in Senegal. Through selection mechanisms, a dog-related RABV isolate
50
51 368 may have acquired the key genetic information to induce host switching, adaptation and viral
52
53 369 spillover into honey-badgers as the honey-badger-related RABV clustered with a dog-related
54
55 370 isolate collected from the same place during the same year. In addition, through the identification
56
57 371 of a quietly similar mongoose-related mutation in a dog-related RABV sequence (Pro-386G-
58
59 372 Leu), our data could worth consider that the Pro-386G-Ser mutation previously identified as

1
2
3 373 specific to mongoose-related RABV (Africa-3 clade) could be also present in dog-related RABV
4
5 374 isolates. Nevertheless, our phylogenetic data has exhibited the impact of purifying selection in
6
7 375 the molecular evolutionary dynamic of the Africa-2 clade RABV circulating in Senegal. As rabid
8
9 376 dogs constitute the main source of transmission to humans, Rabies control in dog populations
10
11 377 through a strong one health cooperation is essential for successful elimination in Senegal by
12
13 378 2030. However, the Rabies spillover in honey-badger exhibits the importance of establishing a
14
15 379 surveillance program in wildlife, to improve control measures and avoid possible future re-
16
17 380 introductions into domestic hosts.

18 381 **Acknowledgments**

19
20
21 382 This research received no specific grant from any funding agency in the public, commercial, or
22
23 383 not-for-profit sectors and was only supported by the Institut Pasteur de Dakar. We acknowledge
24
25 384 colleagues of virology department at Institut Pasteur de Dakar, Senegal for sharing supportive
26
27 385 information necessary for establishment and accomplishment of this study.

28 29 386 **Conflict of interest statement**

30
31 387 The authors declared no competing interest.

32 33 34 388 **References**

- 35
36
37 389 1. World Health Organization (2013). World Health Organization expert consultation on
38 390 rabies. World Health Organization Technical report series, 931, 1–88.
- 39 391 2. Hampson, K., Coudeville, L., Lembo, T., Sambo, M., Kieffer, A., Attlan, M., Barrat, J.,
40 392 Blanton, J. D., Briggs, D. J., Cleaveland, S., Costa, P., Freuling, C. M., Hiby, E., Knopf,
41 393 L., Leanes, F., Meslin, F. X., Metlin, A., Miranda, M. E., Müller, T., Nel, L. H., ...
42 394 Global Alliance for Rabies Control Partners for Rabies Prevention (2015). Estimating the
43 395 global burden of endemic canine rabies. *PLoS neglected tropical diseases*, 9(4),
44 396 e0003709. <https://doi.org/10.1371/journal.pntd.0003709>.
- 45 397 3. World Health Organization (2013). Expert Consultation on Rabies—Second report.
46 398 Geneva, Switzerland: World Health Organisation; 2013; 139 p.
- 47 399 4. Riedel, C., Hennrich, A. A., & Conzelmann, K. K. (2020). Components and Architecture
48 400 of the Rhabdovirus Ribonucleoprotein Complex. *Viruses*, 12(9), 959.
49 401 <https://doi.org/10.3390/v12090959>.
- 50 402 5. Luo, J., Zhang, Y., Zhang, Q., Wu, Y., Zhang, B., Mo, M., Tian, Q., Zhao, J., Mei, M., &
51 403 Guo, X. (2019). The Deoptimization of Rabies Virus Matrix Protein Impacts Viral
52 404 Transcription and Replication. *Viruses*, 12(1), 4. <https://doi.org/10.3390/v12010004>.
- 53
54
55
56
57
58
59
60

- 1
2
3 405 6. Zhang, G., Wang, H., Mahmood, F., & Fu, Z. F. (2013). Rabies virus glycoprotein is an
4 406 important determinant for the induction of innate immune responses and the pathogenic
5 407 mechanisms. *Veterinary microbiology*, 162(2-4), 601–613.
6 408 <https://doi.org/10.1016/j.vetmic.2012.11.031>.
- 7
8 409 7. Li, J., Ertel, A., Portocarrero, C., Barkhouse, D. A., Dietzschold, B., Hooper, D. C., &
9 410 Faber, M. (2012). Postexposure treatment with the live-attenuated rabies virus (RV)
10 411 vaccine TriGAS triggers the clearance of wild-type RV from the Central Nervous System
11 412 (CNS) through the rapid induction of genes relevant to adaptive immunity in CNS
12 413 tissues. *Journal of virology*, 86(6), 3200–3210. <https://doi.org/10.1128/JVI.06699-11>.
- 13
14 414 8. Rupprecht, C. E., Hanlon, C. A., & Hemachudha, T. (2002). Rabies re-examined. *The*
15 415 *Lancet. Infectious diseases*, 2(6), 327–343. [https://doi.org/10.1016/s1473-](https://doi.org/10.1016/s1473-3099(02)00287-6)
16 416 [3099\(02\)00287-6](https://doi.org/10.1016/s1473-3099(02)00287-6)
- 17 417 9. Rupprecht, C. E., Barrett, J., Briggs, D., Cliquet, F., Fooks, A. R., Lumlertdacha, B.,
18 418 Meslin, F. X., Müller, T., Nel, L. H., Schneider, C., Tordo, N., & Wandeler, A. I. (2008).
19 419 Can rabies be eradicated?. *Developments in biologicals*, 131, 95–121.
- 20 420 10. Freire de Carvalho, M., Vigilato, M., Pompei, J. A., Rocha, F., Vokaty, A., Molina-
21 421 Flores, B., Cosivi, O., & Del Rio Vilas, V. J. (2018). Rabies in the Americas: 1998-2014.
22 422 *PLoS neglected tropical diseases*, 12(3), e0006271.
23 423 <https://doi.org/10.1371/journal.pntd.0006271>.
- 24 424 11. Barnard B. J. (1979). The role played by wildlife in the epizootiology of rabies in South
25 425 Africa and South-West Africa. *The Onderstepoort journal of veterinary research*, 46(3),
26 426 155–163.
- 27
28 427 12. Troupin, C., Dacheux, L., Tanguy, M., Sabetta, C., Blanc, H., Bouchier, C., Vignuzzi, M.,
29 428 Duchene, S., Holmes, E. C., & Bourhy, H. (2016). Large-Scale Phylogenomic Analysis
30 429 Reveals the Complex Evolutionary History of Rabies Virus in Multiple Carnivore Hosts.
31 430 *PLoS pathogens*, 12(12), e1006041. <https://doi.org/10.1371/journal.ppat.1006041>.
- 32
33 431 13. Talbi, C., Holmes, E. C., de Benedictis, P., Faye, O., Nakouné, E., Gamatié, D., Diarra,
34 432 A., Elmamy, B. O., Sow, A., Adjogoua, E. V., Sangare, O., Dundon, W. G., Capua, I.,
35 433 Sall, A. A., & Bourhy, H. (2009). Evolutionary history and dynamics of dog rabies virus
36 434 in western and central Africa. *The Journal of general virology*, 90(Pt 4), 783–791.
37 435 <https://doi.org/10.1099/vir.0.007765-0>.
- 38
39 436 14. Davis, P. L., Rambaut, A., Bourhy, H., & Holmes, E. C. (2007). The evolutionary
40 437 dynamics of canid and mongoose rabies virus in Southern Africa. *Archives of virology*,
41 438 152(7), 1251–1258. <https://doi.org/10.1007/s00705-007-0962-9>.
- 42 439 15. OIE. Rabies. 2013. In: *Manual of Diagnostic Tests and Vaccines for Terrestrial Animals*
43 440 2016. Paris: World Organisation for Animal Health. Available from:
44 441 <http://www.oie.int/en/international-standardsetting/terrestrial-manual/access-online/>.
- 45
46 442 16. Faye, M., Faye, O., Diagne, M. M., Fall, G., Weidmann, M., Sembene, M., Sall, A. A., &
47 443 Faye, O. (2018). Full-Genome Characterization and Genetic Evolution of West African
48 444 Isolates of Bagaza Virus. *Viruses*, 10(4), 193. <https://doi.org/10.3390/v10040193>
- 49 445 17. Okonechnikov, K., Golosova, O., Fursov, M., & UGENE team (2012). Unipro UGENE:
50 446 a unified bioinformatics toolkit. *Bioinformatics (Oxford, England)*, 28(8), 1166–1167.
51 447 <https://doi.org/10.1093/bioinformatics/bts091>.
- 52 448 18. Edgar R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high
53 449 throughput. *Nucleic acids research*, 32(5), 1792–1797.
54 450 <https://doi.org/10.1093/nar/gkh340>

- 1
2
3 451 19. Faber, M., Faber, M. L., Papaneri, A., Bette, M., Weihe, E., Dietzschold, B., & Schnell,
4 452 M. J. (2005). A single amino acid change in rabies virus glycoprotein increases virus
5 453 spread and enhances virus pathogenicity. *Journal of virology*, 79(22), 14141–14148.
6 454 <https://doi.org/10.1128/JVI.79.22.14141-14148.2005>.
- 7
8 455 20. Dietzschold, B., Wunner, W. H., Wiktor, T. J., Lopes, A. D., Lafon, M., Smith, C. L., &
9 456 Koprowski, H. (1983). Characterization of an antigenic determinant of the glycoprotein
10 457 that correlates with pathogenicity of rabies virus. *Proceedings of the National Academy*
11 458 *of Sciences of the United States of America*, 80(1), 70–74.
12 459 <https://doi.org/10.1073/pnas.80.1.70>.
- 13
14 460 21. Marston, D. A., McElhinney, L. M., Johnson, N., Müller, T., Conzelmann, K. K., Tordo,
15 461 N., & Fooks, A. R. (2007). Comparative analysis of the full genome sequence of
16 462 European bat lyssavirus type 1 and type 2 with other lyssaviruses and evidence for a
17 463 conserved transcription termination and polyadenylation motif in the G-L 3' non-
18 464 translated region. *The Journal of general virology*, 88(Pt 4), 1302–1314.
19 465 <https://doi.org/10.1099/vir.0.82692-0>.
- 20
21 466 22. Tomar N.R., Kumar R., & Kumar A. (2017). Glycoprotein Based Phylogenetic Analysis
22 467 of Rabies Virus Isolates. *International Journal of Current Microbiology and Applied*
23 468 *Sciences*, 6(7), 1727-1734. doi: <https://doi.org/10.20546/ijcmas.2017.607.208>.
- 24 469 23. Martin, D. P., Murrell, B., Golden, M., Khoosal, A., & Muhire, B. (2015). RDP4:
25 470 Detection and analysis of recombination patterns in virus genomes. *Virus evolution*, 1(1),
26 471 vev003. <https://doi.org/10.1093/ve/vev003>.
- 27 472 24. Pond, S. L., Frost, S. D., & Muse, S. V. (2005). HyPhy: hypothesis testing using
28 473 phylogenies. *Bioinformatics (Oxford, England)*, 21(5), 676–679.
29 474 <https://doi.org/10.1093/bioinformatics/bti079>.
- 30
31 475 25. Rambaut, A., Lam, T. T., Max Carvalho, L., & Pybus, O. G. (2016). Exploring the
32 476 temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen).
33 477 *Virus evolution*, 2(1), vew007. <https://doi.org/10.1093/ve/vew007>.
- 34 478 26. Wickham, H. (2016). *ggplot2: elegant graphics for data analysis*. Springer.
- 35 479 27. Hasegawa, M., Kishino, H., & Yano, T. (1985). Dating of the human-ape splitting by a
36 480 molecular clock of mitochondrial DNA. *Journal of molecular evolution*, 22(2), 160–174.
37 481 <https://doi.org/10.1007/BF02101694>.
- 38
39 482 28. Gill, M. S., Lemey, P., Faria, N. R., Rambaut, A., Shapiro, B., & Suchard, M. A. (2013).
40 483 Improving Bayesian population dynamics inference: a coalescent-based model for
41 484 multiple loci. *Molecular biology and evolution*, 30(3), 713–724.
42 485 <https://doi.org/10.1093/molbev/mss265>.
- 43
44 486 29. Ferreira, M. A. R., Suchard, M. A. (2008). Bayesian analysis of elapsed times in
45 487 continuous-time Markov chains. *The Canadian Journal of Statistics*, 36, 355-368.
46 488 <https://doi.org/10.1002/cjs.5550360302>.
- 47 489 30. Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior
48 490 Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic biology*, 67(5),
49 491 901–904. <https://doi.org/10.1093/sysbio/syy032>.
- 50 492 31. Rambaut A, Drummond AJ. 2010. TreeAnnotator version 1.6.1 [computer program]
51 493 Retrieved from: <http://beast.bio.ed.ac.uk>.
- 52
53 494 32. Zhou, M., Zhou, Z., Kia, G. S., Gnanadurai, C. W., Leyson, C. M., Umoh, J. U., Kwaga,
54 495 J. P., Kazeem, H. M., & Fu, Z. F. (2013). Complete genome sequence of a street rabies
55
56
57
58
59
60

- 1
2
3 496 virus isolated from a dog in Nigeria. *Genome announcements*, 1(1), e00214-12.
4 497 <https://doi.org/10.1128/genomeA.00214-12>.
- 5 498 33. Tricou, V., Berthet, N., Nakouné, E., & Kazanji, M. (2014). Complete genome sequence
6 499 of a rabies virus isolated from a human in central african republic. *Genome*
7 500 *announcements*, 2(3), e00598-14. <https://doi.org/10.1128/genomeA.00598-14>.
- 8 501 34. Mauti S, Léchenne M, Mbilo C, Nel L, Zinsstag J. Rabies. In: Kardjadj M., Diallo A.,
9 502 Lancelot R. (eds) (2019). *Transboundary Animal Diseases in Sahelian Africa and*
10 503 *Connected Regions*. Springer, Cham. 107-119. [https://doi.org/10.1007/978-3-030-25385-](https://doi.org/10.1007/978-3-030-25385-1)
11 504 1.
- 12 505 35. Bingham, J., Hill, F. W., & Matema, R. (1994). Rabies incubation in an African civet
13 506 (*Civettictis civetta*). *The Veterinary record*, 134(20), 528.
14 507 <https://doi.org/10.1136/vr.134.20.528>.
- 15 508 36. Bingham, J., Javangwe, S., Sabeta, C. T., Wandeler, A. I., & Nel, L. H. (2001). Report of
16 509 isolations of unusual lyssaviruses (rabies and Mokola virus) identified retrospectively
17 510 from Zimbabwe. *Journal of the South African Veterinary Association*, 72(2), 92–94.
18 511 <https://doi.org/10.4102/jsava.v72i2.624>.
- 19 512 37. Schnell, M. J., McGettigan, J. P., Wirblich, C., & Papaneri, A. (2010). The cell biology
20 513 of rabies virus: using stealth to reach the brain. *Nature reviews. Microbiology*, 8(1), 51–
21 514 61. <https://doi.org/10.1038/nrmicro2260>.
- 22 515 38. Hemachudha, T., Ugolini, G., Wacharapluesadee, S., Sungkarat, W., Shuangshoti, S., &
23 516 Laothamatas, J. (2013). Human rabies: neuropathogenesis, diagnosis, and management.
24 517 *The Lancet. Neurology*, 12(5), 498–513. [https://doi.org/10.1016/S1474-4422\(13\)70038-](https://doi.org/10.1016/S1474-4422(13)70038-3)
25 518 3.
- 26 519 39. Seif, I., Coulon, P., Rollin, P. E., & Flamand, A. (1985). Rabies virulence: effect on
27 520 pathogenicity and sequence characterization of rabies virus mutations affecting antigenic
28 521 site III of the glycoprotein. *Journal of virology*, 53(3), 926–934.
29 522 <https://doi.org/10.1128/JVI.53.3.926-934.1985>.
- 30 523 40. Yamada, K., Noguchi, K., Nonaka, D., Morita, M., Yasuda, A., Kawazato, H., &
31 524 Nishizono, A. (2013). Addition of a single N-glycan to street rabies virus glycoprotein
32 525 enhances virus production. *The Journal of general virology*, 94(Pt 2), 270–275.
33 526 <https://doi.org/10.1099/vir.0.047852-0>.
- 34 527 41. Deviatkin, A. A., & Lukashev, A. N. (2018). Recombination in the rabies virus and other
35 528 lyssaviruses. *Infection, genetics and evolution: journal of molecular epidemiology and*
36 529 *evolutionary genetics in infectious diseases*, 60, 97–102.
37 530 <https://doi.org/10.1016/j.meegid.2018.02.026>.
- 38 531 42. Ding, N. Z., Xu, D. S., Sun, Y. Y., He, H. B., & He, C. Q. (2017). A permanent host shift
39 532 of rabies virus from Chiroptera to Carnivora associated with recombination. *Scientific*
40 533 *reports*, 7(1), 289. <https://doi.org/10.1038/s41598-017-00395-2>.
- 41 534 43. Mollentze, N., Biek, R., & Streicker, D. G. (2014). The role of viral evolution in rabies
42 535 host shifts and emergence. *Current opinion in virology*, 8, 68–72.
43 536 <https://doi.org/10.1016/j.coviro.2014.07.004>.
- 44 537 44. Randall, D. A., Williams, S. D., Kuzmin, I. V., Rupprecht, C. E., Tallents, L. A., Tefera,
45 538 Z., Argaw, K., Shiferaw, F., Knobel, D. L., Sillero-Zubiri, C., & Laurenson, M. K.
46 539 (2004). Rabies in endangered Ethiopian wolves. *Emerging infectious diseases*, 10(12),
47 540 2214–2217. <https://doi.org/10.3201/eid1012.040080>.

- 1
2
3 541 45. Streicker, D. G., Altizer, S. M., Velasco-Villa, A., & Rupprecht, C. E. (2012). Variable
4 542 evolutionary routes to host establishment across repeated rabies virus host shifts among
5 543 bats. *Proceedings of the National Academy of Sciences of the United States of America*,
6 544 109(48), 19715–19720. <https://doi.org/10.1073/pnas.1203456109>.
- 7 545 46. Katz, I. S., Fuoco, N. L., Chaves, L. B., Rodrigues, A. C., Ribeiro, O. G., Scheffer, K. C.,
8 546 & Asano, K. M. (2016). Delayed progression of rabies transmitted by a vampire bat.
9 547 *Archives of virology*, 161(9), 2561–2566. <https://doi.org/10.1007/s00705-016-2927-3>.
- 10 548 47. Chiou, H. Y., Hsieh, C. H., Jeng, C. R., Chan, F. T., Wang, H. Y., & Pang, V. F. (2014).
11 549 Molecular characterization of cryptically circulating rabies virus from ferret badgers,
12 550 Taiwan. *Emerging infectious diseases*, 20(5), 790–798.
13 551 <https://doi.org/10.3201/eid2005.131389>.
- 14 552 48. Bourhy, H., Reynes, J. M., Dunham, E. J., Dacheux, L., Larrous, F., Huong, V., Xu, G.,
15 553 Yan, J., Miranda, M., & Holmes, E. C. (2008). The origin and phylogeography of dog
16 554 rabies virus. *The Journal of general virology*, 89(Pt 11), 2673–2681.
17 555 <https://doi.org/10.1099/vir.0.2008/003913-0>.
- 18 556 49. Fèvre, E. M., Kaboyo, R. W., Persson, V., Edelsten, M., Coleman, P. G., & Cleaveland,
19 557 S. (2005). The epidemiology of animal bite injuries in Uganda and projections of the
20 558 burden of rabies. *Tropical medicine & international health : TM & IH*, 10(8), 790–798.
21 559 <https://doi.org/10.1111/j.1365-3156.2005.01447.x>.
- 22 560
23 561
24 562
25 563

26
27
28 **Table 1:** List of rabies virus full genome sequences from Senegalese's isolates from 2001 to
29 2015.
30 563

Isolate ID	Collection year	Origin	Host	GenBank Accession Number
SH155966	2001	Dakar	<i>Homo sapiens</i>	MH514968
SH177846	2005	Dakar	<i>Homo sapiens</i>	MH514969
SA217694	2011	Linguère	<i>Canis lupus familiaris</i>	MH514970
SA206776	2010	Dakar	<i>Canis lupus familiaris</i>	MH514971
SA194858	2008	Dakar	<i>Canis lupus familiaris</i>	MH514972
SA212203	2011	Dahra	<i>Mellivora capensis</i>	MH514973
SA217750	2011	Dahra	<i>Canis lupus familiaris</i>	MH514974
SA252913	2013	Diamniadio	<i>Canis lupus familiaris</i>	MH514975
SA204014	2010	Dakar	<i>Canis lupus familiaris</i>	MH514976
SA262037	2013	Dakar	<i>Canis lupus familiaris</i>	MH514977
SA262503	2014	Dakar	<i>Canis lupus familiaris</i>	MH514978
SA262518	2014	Dakar	<i>Canis lupus familiaris</i>	MH514979

SA252888	2013	Dakar	<i>Canis lupus familiaris</i>	MH514980
SA267115	2014	Dakar	<i>Canis lupus familiaris</i>	MH514981
SH189343	2007	Dakar	<i>Homo sapiens</i>	MH514982
SA217695	2011	Dahra	<i>Canis lupus familiaris</i>	MH514983
SH218152	2011	Dakar	<i>Homo sapiens</i>	MH514984
SA272282	2015	Dakar	<i>Canis lupus familiaris</i>	MH514985

564

565 **Table 2:** Amino acid substitutions in genome of the three most divergent Senegalese rabies
 566 virus sequences.

Protein and codon position	Honey-bagder-related RABV	Dog-related RABV	Dog-related RABV
	SA212203_Dahra_SN_2011	SA252888_Dakar_SN_2013	SA267115_Dakar_SN_2014
Glycoprotein			
25			Thr => Lys
118			Asn => Asp
146			Thr => Ile
150		Ser => Cys	
197			Thr => Cys
233			Gly => Ser
248		Gly => Ala	
317			Phe => Ser
321	Ser => Asp		

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

338 Asn => Asp

339 Lys => Glu

347 His => Trp

352 Arg => Gln

356 Glu => Ala

362 Gly => Ala

367 Gly => Asp

368 Gly => Arg

370 Cys => Arg

373 His => Tyr

386 Pro => Leuⁿ

395

415 Met => Gly Met => Thr

417 Pro => Ser

496 Thr => Ala

Polymerase

622 Trp => Arg

625 His => Gln

1220 Leu => Pro

1380	Gly => Trp
1391	Leu => Ser
1423	Arg => Pro
1427	Ser => Leu
1444	Asp => His
1600	Ser => Arg
1633	Ser => Leu

Mutations were assessed with reference to the previously available complete genome sequences from Senegal (KX148238-39).

[¶] Substitution of the amino acid mutation Pro-386G-Ser previously identified as specific to mongoose-related RABV (Africa-3 clade) by Troupin C, et al. 2016 (12).

567

568

569 **Table 3:** Conservation of amino acid residues of virulence across the glycoprotein of Senegalese
570 rabies virus isolates.

Amino acid residues	Position	Motif's conservation	Replaced by this consensus sequence on these RABV isolates
Arginine (R)	333	NO	Alanine (<u>A</u>) on all the Senegalese isolates
Asparagine (N)	194	NO	Tyrosine (<u>Y</u>) on all the Senegalese isolates located at position 338 NK <u>A</u> on the isolate SA194858_Dakar_SN_2008;
Glycosylation site NKF	319	NO	NE <u>A</u> on the isolate SA252888_Dakar_SN_2013; D <u>K</u> T on the isolate SA212203_Dahra_SN_2011
EMQSSLLQQH	394-403	YES	ETQSSLLQQ <u>L</u> on the isolate SA267115_Dakar_SN_2014;

			EMQSSLLQQL on the isolate SA252888_Dakar_SN_2013
NHDYTIWMPE	191-200	YES	NHDYTI <u>C</u> MPE on the isolate SA267115SEN_Dakar_SN_2014
HNPYPDYHWL	132-141	YES	HNS <u>Y</u> PDYHWL on the isolate SA252913_Diamniadio_SN_2013
AETYTNFVGY	87-96	YES	Similar on all the Senegalese isolates
TTFKRKHFRP	99-108	YES	TTFRRKHFRP on all the Senegalese isolates
DIHHLSCPNN	37-46	YES	Similar on all the Senegalese isolates
KWCSPDQLVN	269-278	YES	KWC <u>P</u> PDQLVN on the isolates SH155996_Dakar_SN_2001 and SH177846_Dakar_SN_2005; RW <u>C</u> PDQLVN on the isolate SA252888_Dakar_SN_2013; KW <u>C</u> <u>P</u> SDQLVN on all the other Senegalese isolates
WKMAGDPRYE	119-128	YES	WKMAG <u>G</u> PRYE on the isolate SA212203_Dahra_SN_2011
CGFVDERGLY	226-235	YES	CGFVDER <u>S</u> LY on the isolate SA267115_Dakar_SN_2014
TLMEADAHYK	341-350	YES	<u>A</u> LMEAH <u>A</u> HYK on the isolates SA252888_Dakar_SN_2013 and SA194858_Dakar_SN_2008; TLMEADA <u>W</u> YK on the isolate SA212203_Dahra_SN_2011

The residues **Arg333** and **Asn194** were previously described by **Dietzschold B et al, 1983 (20)**. The N-glycosylation motif **NKT** was previously described by **Marston DA et al, 2007 (21)**. The ten remaining RABV amino acid motifs were analyzed as described by and **Tomar NR et al, 2017 (22)**.

Positions with amino acid different on the glycoprotein of Senegalese RABV are highlighted in bold and black for conservative mutations and in bold and red for non-conservative mutations.

A motif was considered as non-conserved when it presents more than three non-conservative mutations in more than three RABV isolates.

571

572 **Table 4:** Episodes of positive diversifying selection among the proteins of Senegalese rabies
573 virus isolates.

Protein	Number of sites detected by method				Positive selection
	SLAC (P value ≤ 0.1)	FUBAR (PP ≥ 0.9)	MEME (P value ≤ 0.1)	aBSREL (P value < 0.05)	
N	58/540	269/540	0/540	0/102	YES
	0/540	1/540	1/540	0/102	
P	23/297	43/297	0/297	0/55	YES
	0/297	1/297	1/297	0/55	
M	12/202	39/202	0/202	0/55	YES
	0/202	1/202	1/202	0/55	
G	55/524	269/524	0/524	0/75	YES
	0/524	0/524	6/524	1/75	
L	110/2127	813/2127	0/2127	0/67	NO
	0/2127	0/2127	8/2127	0/67	

For each protein, the top row represents sites under negative selection ($dN/dS < 1$), and the bottom row represents site that are under positive selection ($dN/dS > 1$).

Pervasive diversifying selection at posterior probability (PP) ≥ 0.9 with FUBAR model

Episodic diversifying selection at 0.1 significance level with SLAC and MEME models (P value $p \leq 0.1$)

Episodic diversifying selection at P value $p \leq 0.05$ with aBSREL model.

574

1
2
3 **575 Figure legends**
4

5 **576 Figure 1:** Multiple sequence alignment of Senegalese RABV G protein sequences at antigenic
6 sites I (residue 231), II (residues 34–42, 198–200), III (residues 330–338) and IV (residue 342).
7
8 578 10 vaccine strain G protein sequences were also included for medical countermeasure
9
10 579 comparisons.
11

12 580

13 **581 Figure 2:** (A) A root-to-tip regression analysis (top-left) using a maximum-likelihood estimated
14 tree (right) with 37 RABV complete genome sequences. Red circles denote non-Senegalese
15 582 isolates while yellow triangles represent Senegalese isolates. (B) A maximum-likelihood tree
16 583 estimating phylogenetic relations in West African sequences. Tree branches are scaled by
17 584 substitutions per site. Tree nodes show bootstrap support values when > 700 . Yellow triangles
18 585 denote Senegalese isolates. Tree tip labels include isolate/strain, country and/or city of isolation,
19 586 and collection year. Hosts where RABV was isolated from are also shown for Senegalese
20 587 sequences.
21 588
22
23
24
25
26
27

28 589

29
30 **590 Figure 3:** A Bayesian-inferred maximum-clade credibility tree estimated using complete RABV
31 591 genomes, date of isolation, and geographical location. Phylogeographic estimation of migration
32 592 events into Senegal are shown in the top-left corner. Estimated time to the most recent common
33 593 ancestor (tMRCA) and the 95% highest posterior densities (HPD) are shown at key tree nodes.
34 594 Posterior probabilities > 0.7 are shown. Tree branches are colored by country and scaled by
35 595 substitutions per site per year. Tree tips labels include the strain/isolate, country and year of
36 596 isolation. Tree tip nodes with Senegalese taxon are colored in light purple.
37
38
39
40
41
42
43

44 597

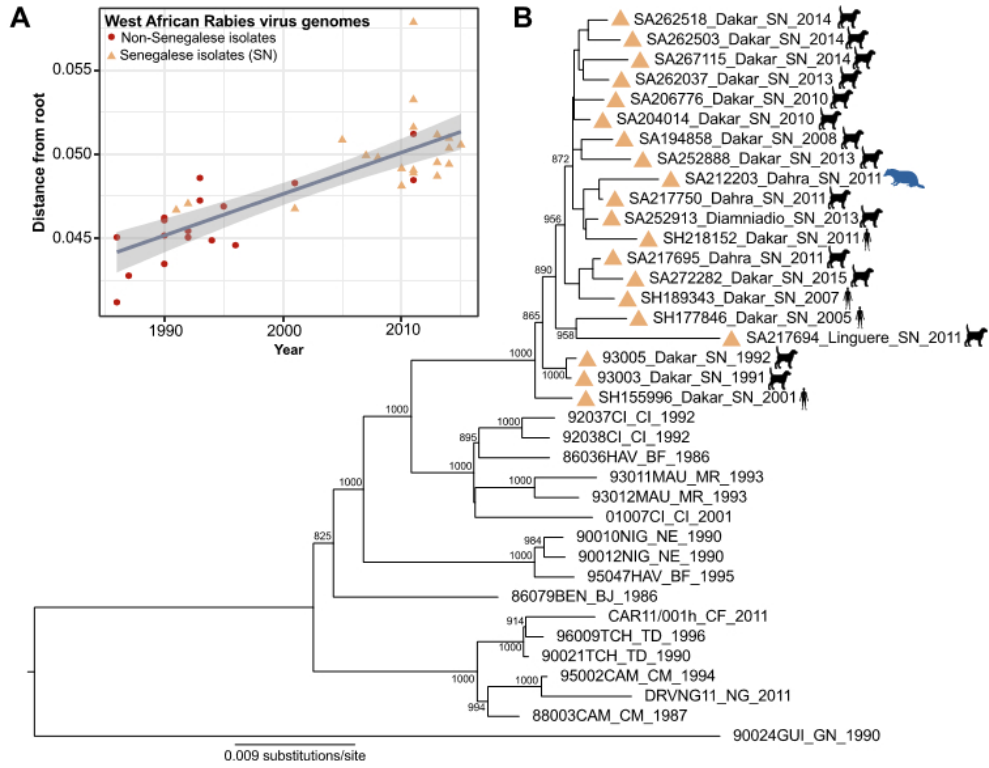
45
46 598
47
48
49
50
51
52
53
54
55
56
57
58
59
60



Appendix Figure 1: The Senegalese map showing the localities of Dahra and Linguère. The red stars in the map indicate the geographical position of the localities of Dahra and Linguère in the Louga region, Senegal.

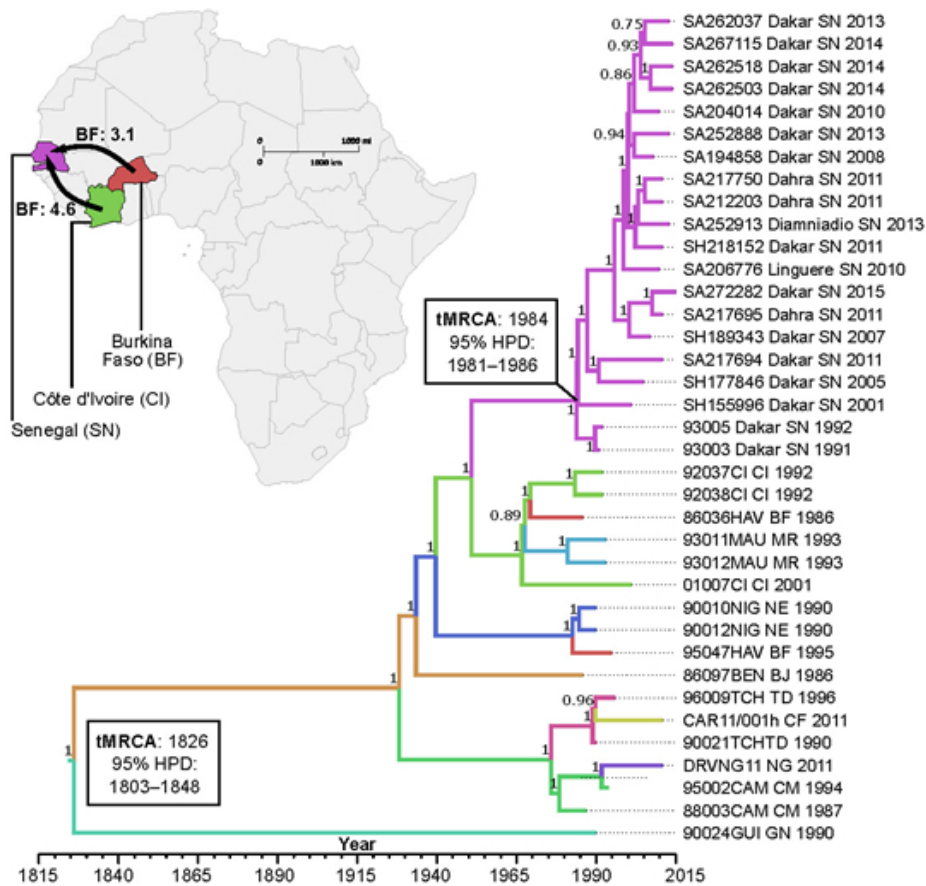
Isolate	Antigenic Sites					
	34-42	198-200	231	264	330-338	342
SH155996_Dakar_SN_2001	SPIDIHHL	MPE	E	T	FGKAYTIFN	M
SH177846_Dakar_SN_2005	SPIDIHHL	MPE	E	T	FGKAYTIFN	M
SA217694_Dakar_SN_2011	SPIDIHHL	MPE	E	T	FGKAYTIFN	M
SA206776_Linguere_SN_2010	SPIDIHHL	MPE	E	T	FGKAYTIFN	M
SA194858_Dakar_SN_2008	SPIDIHHL	MPE	E	T	FGKAYTIFN	M
SA212203_Dahra_SN_2011	SPIDIHHL	MPE	E	T	FGKAYTIFN	M
SA217750_Dahra_SN_2011	SPIDIHHL	MPE	E	T	FGKAYTIFN	M
SH218152_Dakar_SN_2011	SPIDIHHL	MPE	E	T	FGKAYTIFN	M
SA252913_Diamniadio_SN_2013	SPIDIHHL	MPE	E	T	FGKAYTIFN	M
SA204014_Dakar_SN_2010	SPIDIHHL	MPE	E	T	FGKAYTIFN	M
SA217695_Dahra_SN_2011	SPIDIHHL	MPE	E	T	FGKAYTIFN	M
SA262037_Dakar_SN_2013	SPIDIHHL	MPE	E	T	FGKAYTIFN	M
SA262518_Dakar_SN_2014	SPIDIHHL	MPE	E	T	FGKAYTIFN	M
SA252888_Dakar_SN_2013	SPIDIHHL	MPE	E	T	FGKAYTIFN	M
SA267115_Dakar_SN_2014	SPIDIHHL	MPE	E	T	FGKAYTIFN	M
SA262503_Dakar_SN_2014	SPIDIHHL	MPE	E	T	FGKAYTIFN	M
SH189343_Dakar_SN_2007	SPIDIHHL	MPE	E	T	FGKAYTIFN	M
SA272282_Dakar_SN_2015	SPIDIHHL	MPE	E	T	FGKAYTIFN	M
93005_Dakar_SN_1992	SPIDIHHL	MPE	E	T	FGKAYTIFN	M
93003_Dakar_SN_1991	SPIDIHHL	MPE	E	T	FGKAYTIFN	M
Pasteur	SPIDIHHL	MPE	E	T	FGKAYTIFN	M
Nishigahara	SPIDIHHL	MPE	E	T	FGKAYTIFN	M
RC-HL	SPIDIHHL	MPE	E	T	FGKAYTIFN	M
NI-CE	SPIDIHHL	MPE	E	T	FGKAYTIFN	M
SRV9	SPIDIHHL	MPE	E	T	FGKAYTIFN	M
Flury-LEP	SPIDIHHL	MPE	E	T	FGKAYTIFN	M
ERA	SPIDIHHL	MPE	E	T	FGKAYTIFN	M
SAG2	SPIDIHHL	MPE	E	T	FGKAYTIFN	M
RV-97	SPIDIHHL	MPE	E	T	FGKAYTIFN	M
CTN181	SPIDIHHL	MPE	E	T	FGKAYTIFN	M

154x131mm (96 x 96 DPI)



202x156mm (96 x 96 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



154x138mm (96 x 96 DPI)

Appendix Table 1: Description of primer sets used for viral sequencing of rabies virus isolates.

Primer name	3'-5' Sequence	Direction	Location in viral genome	Melting temperature (°C)
SN3race243R	AGCTTGGCAGCATTTCATGCCTG	Reverse	Nucleoprotein	53
SN3race298R	TGCTGCTGCCAAGTAGGAACAT	Reverse	Nucleoprotein	53
N127	ATGTAACACCTCTACAATGG	Foward	Nucleoprotein	56
N8m	CAGTCTCYTCNGCCATCTC	Reverse	Nucleoprotein	56
SNRV1342F	TATGKTCAGTCAGTTCC	Foward	Nucleoprotein	55
SNRV2939R	TCRTCCCAAGTGATCTCY	Reverse	Matrix	55
SNRV2745F	ATATTCYGGAAAYCACAGRAT	Foward	Matrix	53
SNRV4280R	TGAGACGTCTRAAACTCACTG	Reverse	Glycoprotein	53
M220	TGGTGTATCAACATGRAYTC	Foward	Matrix	53
SNG5444R	GGTCATCATAGACCTCYC	Reverse	Glycoprotein	53
G4836-S3	GGRARRGTYATATCTTCNTGGGA	Foward	Glycoprotein	54
PV08	GGTCTGATCTRTCWGARYAATA	Reverse	polymerase L	54
L7386-AS3	CTRTC BGARTARTADAYCCANGACT T	Reverse	polymerase L	54
SNRV5318F	GGGCTGGATCATCTATGCTT	Foward	Glycoprotein	53
SNRV6304R	TTGATGACCTCGTAGCCTGA	Reverse	polymerase L	53
SNRV6281F	TGYGGAAAYTCCGGCTAT	Foward	polymerase L	54
SNRV8504R	GCTCRCTGAGAAATCGRG	Reverse	polymerase L	54
Taq3long	ATGAGAAGTGGAAYAAAYCATCA	Foward	polymerase L	54
L9633-AS3	TTGCYRTATATGTTGACAGG	Reverse	polymerase L	54
SNRV8401F	TTCAGAGTTYAGAGAGGCRAT	Foward	polymerase L	53
SNRV9648R	ATGTTAACAGGGAAGATTGTT	Reverse	polymerase L	53
SNRV9263F	ATGTTYCAGCCATTGATGCTT	Foward	polymerase L	54
SNRV10590R	TGAAYACAAGCTTRGCATCYG	Reverse	polymerase L	54
SNRV10331F	TGCTCTGCTCAACAGGTT	Foward	polymerase L	53
SNRV11932R	ACGCTTAACAAATAAACAACA	Reverse	5' trailer	53
SN5race11566F	GGTCTGGTGACACCCCGGTCTTCA	Foward	polymerase L	55
SN5race11615F	GAGTCTGTCATCTCACTGGATCA	Foward	polymerase L	55

Appendix Table 2: List of Rabies viruses full genomes from West and Central African countries and available in GenBank.

Isolate	Collection year	Isolation place	Host	GenBank Accession
93005	1992	SN	<i>Canis lupus familiaris</i>	KX148239
93003	1991	SN	<i>Canis lupus familiaris</i>	KX148238
92037CI	1992	CI	<i>Canis lupus familiaris</i>	KX148232
92038CI	1992	CI	<i>Canis lupus familiaris</i>	KX148233
01007CI	2001	CI	<i>Canis lupus familiaris</i>	KX148235
86036HAV	1986	BF	<i>Canis lupus familiaris</i>	KX148234
95047HAV	1995	BF	<i>Canis lupus familiaris</i>	KX148230
93011MAU	1993	MAU	<i>Canis lupus familiaris</i>	KX148236
93012MAU	1993	MAU	<i>Canis lupus familiaris</i>	KX148237
90010NIG	1990	NE	<i>Canis lupus familiaris</i>	KX148231
90012NIG	1990	NE	<i>Canis lupus familiaris</i>	KX148229
DRV_NG11	2011	NG	<i>Canis lupus familiaris</i>	KC196743
86097BEN	1986	BJ	<i>Felis catus</i>	KX148107
90021TCH	1990	TD	<i>Canis lupus familiaris</i>	KX148240
96009TCH	1996	TD	<i>Canis lupus familiaris</i>	KX148241
CAR_11_00 1	2011	CF	<i>Homo sapiens</i>	KF977826
88003CAM	1987	CM	<i>Canis lupus familiaris</i>	KX148243
95002CAM	1994	CM	<i>Canis lupus familiaris</i>	KX148242
90024GUI	1990	GN	<i>Canis lupus familiaris</i>	KX148244

Abbreviations: SN, Senegal. CI, Côte d'Ivoire. BF, Burkina Faso. MAU, Mauritania. NE, Niger. NG, Nigeria. BJ, Benin. TD, Chad. CF, Central African Republic. CM, Cameroon. GN, Guinea

1 **Appendix Table 3:** Bayes factors >3 and posterior probabilities generated from Spread3.

FROM	TO	BAYES_FACTOR	POSTERIOR PROBABILITY
TD	CF	158.0	0.94
CM	NG	123.5	0.93
CM	TD	26.9	0.74
NE	BF	22.1	0.70
CI	MR	11.3	0.55
CI	BF	10.3	0.53
BF	NE	4.7	0.34
CI	SN	4.6	0.33
BF	CI	4.5	0.33
BF	MR	4.4	0.32
BN	CM	3.9	0.29
MR	CI	3.6	0.28
BN	NE	3.5	0.27
BF	SN	3.1	0.25

Abbreviations: SN, Senegal. CI, Côte d'Ivoire. BF, Burkina Faso. MAU, Mauritania. NE, Niger.

NG, Nigeria. BJ, Benin. TD, Chad. CF, Central African Republic. CM, Cameroon. GN, Guinea

2

3

4