

1 **Phylogeographic analysis and identification of factors impacting**
2 **the diffusion of Foot-and-Mouth disease virus in Africa**

3 Florian Duchatel¹, Prof. Mark Bronsvoort¹, Dr. Samantha Lycett^{1*}

4 ¹Roslin Institute, University of Edinburgh, Easter Bush Campus, Midlothian EH25 9RG, UK

5

6 *Corresponding author: Samantha Lycett, The Roslin Institute, University of Edinburgh

7 Email: Samantha.lycett@ed.ac.uk

8 Phone: +44 (0) 131 651 9232

9

10 **ABSTRACT**

11 Foot and mouth disease (FMD) is endemic in sub-Saharan Africa and can lead to important and
12 continuous economic losses for affected countries. Due to the complexity of the disease epidemiology
13 and the lack of data there is a need to use inferential computational approaches to fill the gaps in our
14 understanding of the circulation of FMD virus on this continent. Using a phylogeographic approach we
15 reconstructed the circulation of FMD virus serotypes A, O and SAT2 in Africa and evaluated the
16 influence of potential environmental and anthropological predictors of virus diffusion. Our results show
17 that over the last hundred year the continental circulation of the tree serotypes was mainly driven by
18 livestock trade. Whilst our analyses show that the serotypes A and O were introduced in Africa through
19 livestock trades, the SAT2 serotype probably originates from African wildlife population. The
20 circulation of serotype O in eastern Africa is impacted by both indirect transmission through
21 persistence in the environment and anthropological activities such as cattle movements.

22

23 Foot and mouth disease (FMD) affects more than 70 species of cloven-hoofed animals ¹. The disease
24 is characterised by the development of vesicles in and around the mouth, on the feet and other sites
25 of the skin ¹. The causal agent is a positive-sense, single-stranded RNA virus of the *Picornaviridae*
26 family ² called foot-and-mouth disease virus (FMDV). Its genome encodes the information for 4
27 structural proteins (VP1-4) and 8 non-structural proteins (7 proteases and one RNA polymerase).
28 Based on the level of cross protection between strains, the virus can be divided into seven serotypes,
29 O, A, C, Southern African Territories [SAT] 1, SAT 2, SAT 3 and Asia 1 ^{3,4}, which are clinically
30 indistinguishable from each other but which have different epidemiologies. The hosts that are
31 considered to play an active role in these epidemiologies are cattle, buffaloes, pigs, sheep and goats
32 ⁵.

33 FMD susceptibility varies according to the host and strain of FMDV involved. The severity of the
34 infection depends of the amount of virus inoculated, the serotype, the host species and the individual
35 immunity ⁶. The commonest route of infection for a new host is by direct contact with an infected
36 animal ^{1,7}. The infection may also occur indirectly by contacts with contaminated surfaces or products,
37 such as infected personnel, vehicle or fomites ⁵. Movement of animals and animal products are
38 considered to play an important role in the disease circulation in endemic areas and are considered
39 the main factors for FMDV transboundary spread ⁸.

40 FMD has been eradicated in many high income countries but is still endemic in numerous low and
41 middle income countries (LMICs) ⁹ particularly in Africa and South and East Asia. Although FMD has
42 a low mortality rate in adult animals it causes significant productivity losses that may lead to important
43 and continuous economic losses for farmers and impact countries trading ability at a national level ⁸.
44 Although work has been done to understand the impact of FMDV in large scale dairy farms in LMICs
45 ¹⁰ there is still a lack of data to quantify its impact more broadly on the economy of endemically
46 infected countries ¹¹.

47 FMDV is endemic in most of sub-Saharan Africa with an epidemiology considered to be more
48 complex than in other regions of the world due to multiple serotype and wildlife reservoirs ¹².
49 However, due to a general lack of surveillance and animal traceability, very few statistics on disease
50 incidence and circulation exist for Africa. Although there are a few studies on animal trade and
51 seasonal migration of nomadic and pastoralist herds in sub-Saharan African ^{8,12-14} we need analytical

52 approaches that use existing data to improve our understanding of the circulation of FMDVs in this
53 part of the world.

54 Many wildlife species can be infected by FMDVs in Africa⁶ but amongst all these potential hosts, only
55 the Cape buffalo (*Syncerus caffer*) and impala have been implicated in the transmission of FMDV to
56 domestic cattle¹⁵. Even though the Cape buffalo is suspected to be the primary reservoir and the
57 main source of SAT serotypes in Southern Africa¹⁶, its role as a viral source for livestock epidemics
58 for the FMDV O, A and C serotypes elsewhere in Africa is still unclear^{11,13,17,18}, and might be
59 inexistent¹⁹.

60 It has already been observed that the spatio-temporal occurrence and circulation of FMDV in Africa is
61 mainly affected by human activities though domestic animal movements²⁰⁻²³. However, several
62 environmental characteristics and attributes such as the landscape, vegetation, natural barriers to
63 animal movements such as roads, rivers or mountains have the potential to influence the dynamics
64 and circulation of FMD²⁴⁻²⁶.

65 Because FMDV are single stranded RNA viruses and lack a proof-reading mechanism for their
66 genome replication they are subject to a high evolutionary mutation rate⁴. The history of these
67 mutations can provide information on the ecological processes and population events that shaped the
68 virus evolution even if not directly observed. These processes, along with other evolutionary
69 parameters, can be modelled while reconstructing the phylogenetic trees^{27,28}. Furthermore, by
70 combining genetic data and spatial information, phylogeographic tree reconstruction can be used to
71 estimate the unobserved geographical circulation of a pathogen²⁹. Virus movements can be modelled
72 as discrete transmission events between the sampled locations³⁰ or as a continuous process using
73 different random walk diffusion models³¹. Recently both discrete and continuous approaches have
74 been extended to test and quantify the contribution of potential environmental and anthropological
75 parameters (predictors of viral diffusion) that might influence the spread and circulation of the studied
76 pathogen^{32,33}.

77 The aim of this paper is to gain a better understanding of the circulation of FMDV in Africa, comparing
78 discrete and continuous approaches^{32,34}. A detailed discrete phylogeographical analysis of serotypes
79 A, O and SAT2 sequences was performed, and the influence of 13 potential environmental and

80 anthropological predictors of virus diffusion were quantified and tested using both discrete and

81 continuous approaches.

82

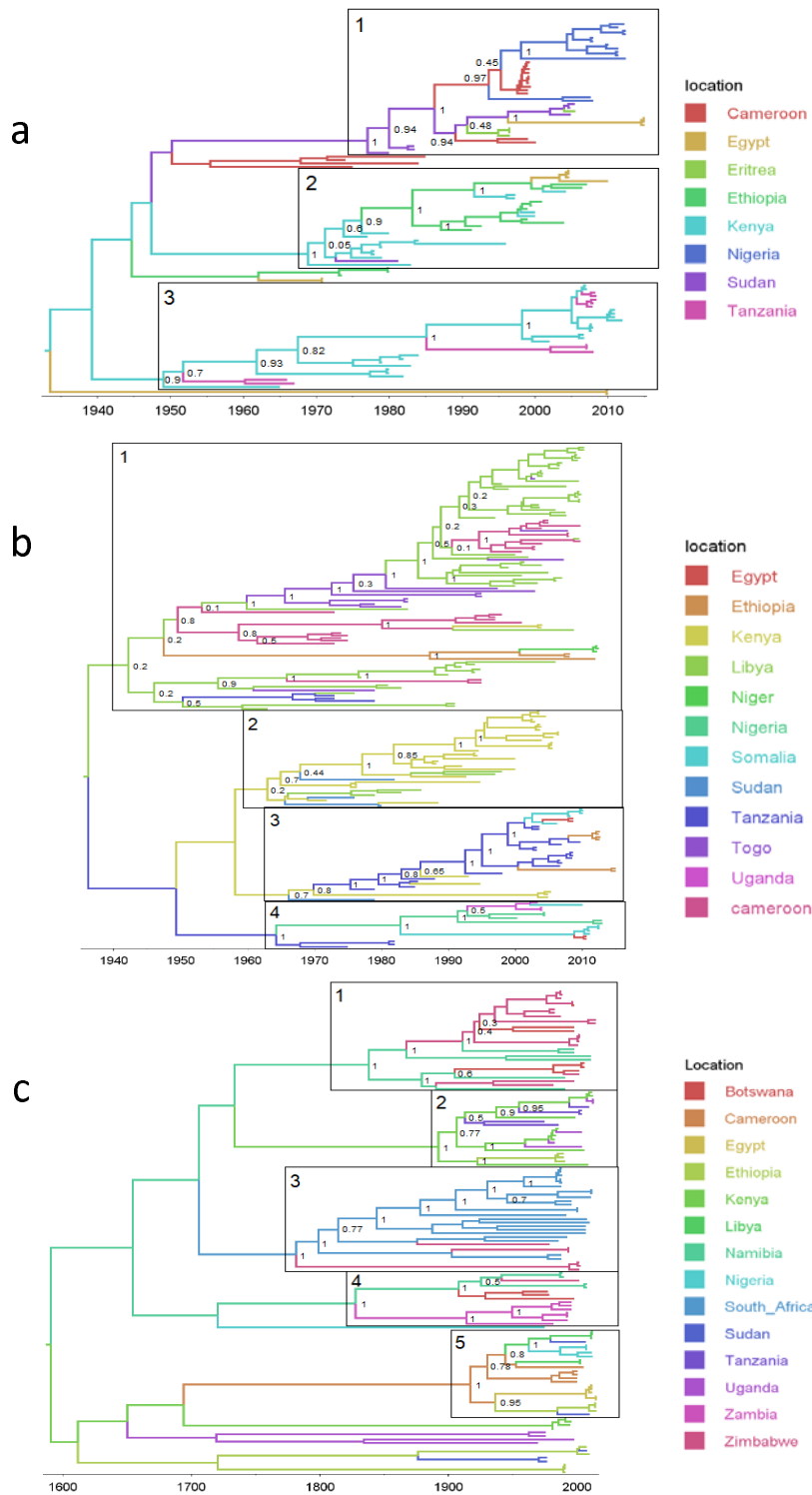
83 **RESULTS**

84 **Discrete phylogenetic analysis**

85 Evolutionary parameters estimation

86

87 Using a relaxed log-normal clock and a skygrid population model for the serotype A and O; and a
88 constant clock and skygrid population model for the serotype SAT2 we estimated evolutionary
89 parameters and reconstructed the phylogenetic trees (see Figure 1). Overall, we observed a mean
90 mutation rate of 4.67×10^{-3} substitutions per site per year and 3.69×10^{-3} for the serotypes A and O
91 respectively. We also estimated a significantly slower mutation rate of 1.1×10^{-3} for the serotype SAT2
92 (see supp. table 1).



93

94 **Figure 1: Bayesian MCC time scaled discrete phylogeographic tree for the three studied serotypes. a.**
 95 **Bayesian phylogeographic tree for serotype A using 107 VP1 sequences. b. Bayesian phylogeographic**
 96 **tree for serotype O using 192 VP1 sequences. c. Bayesian phylogeographic tree for serotype SAT2 using**
 97 **135 VP1 sequences. The phylogeny branches are coloured according to their descendent nodes location**
 98 **with the key for colours shown on the right. The main clades for each of the studied serotypes are**
 99 **identified on the phylogeographic trees. The nodes of the isolated clades are annotated with their**
 100 **posterior probabilities.**

101 Phylogeographic tree reconstruction for serotype A

102 The reconstructed phylogeographic tree of the African serotype A viruses has a time to most recent
103 common ancestor (TMRCA) of around 1926 (1889.6 – 1950 95%HPD), with geographic origin in the
104 eastern part of Africa and high posterior probabilities for Kenya (49.83 %) and Ethiopia (35.95 %)
105 (see fig. 1a). For serotype A there is no clear clade separation between the western and eastern sides
106 of Africa, as the first isolated clade combines all the western African sequences as well as sequences
107 from Sudan, Ethiopia and Egypt. Although a few transmissions events are observable between the
108 two sides of Africa, all of them involve Sudan as a link between them.

109 Phylogeographic tree reconstruction for serotype O

110 The TMRCA of the African serotype O is estimated to be 1937 (1921 – 1952 95%HPD) and located in
111 the eastern part of Africa with high posterior probabilities for Kenya (61.49 %), Sudan (17.15 %) and
112 Uganda (11.42 %) (see fig.1b). The reconstructed phylogeographical tree is composed of four
113 geographically defined main clades. The first clade is almost entirely composed of Kenyan, Tanzanian
114 and Ugandans sequences with only a few transmissions to other countries. The second clade is
115 mostly situated in Ethiopia with few transitions to Kenya and Somalia. The third clade is centred in
116 Sudan with incursions into Nigeria, Cameroon, Egypt and Ethiopia. The fourth clade is centred in
117 West and Central African countries (Cameroon, Nigeria, Niger and Togo) and seems to originate from
118 Sudan. Overall, we can see that the situation for the serotype O is quite similar to that of serotype A
119 with only few observed transmissions between the Eastern and western sides of Africa, with Sudan
120 acting as a link between the two sides of Africa.

121 Phylogeographic tree reconstruction for serotype SAT2

122 The diversity for serotype SAT2 viruses is much greater than for serotypes A and O, and the TMRCA
123 is much older, estimated as 1583 (1722 – 1440 95% HPD). Due to these long timescales and low
124 posterior probabilities on the location it is difficult to estimate an origin location (see fig.1c).

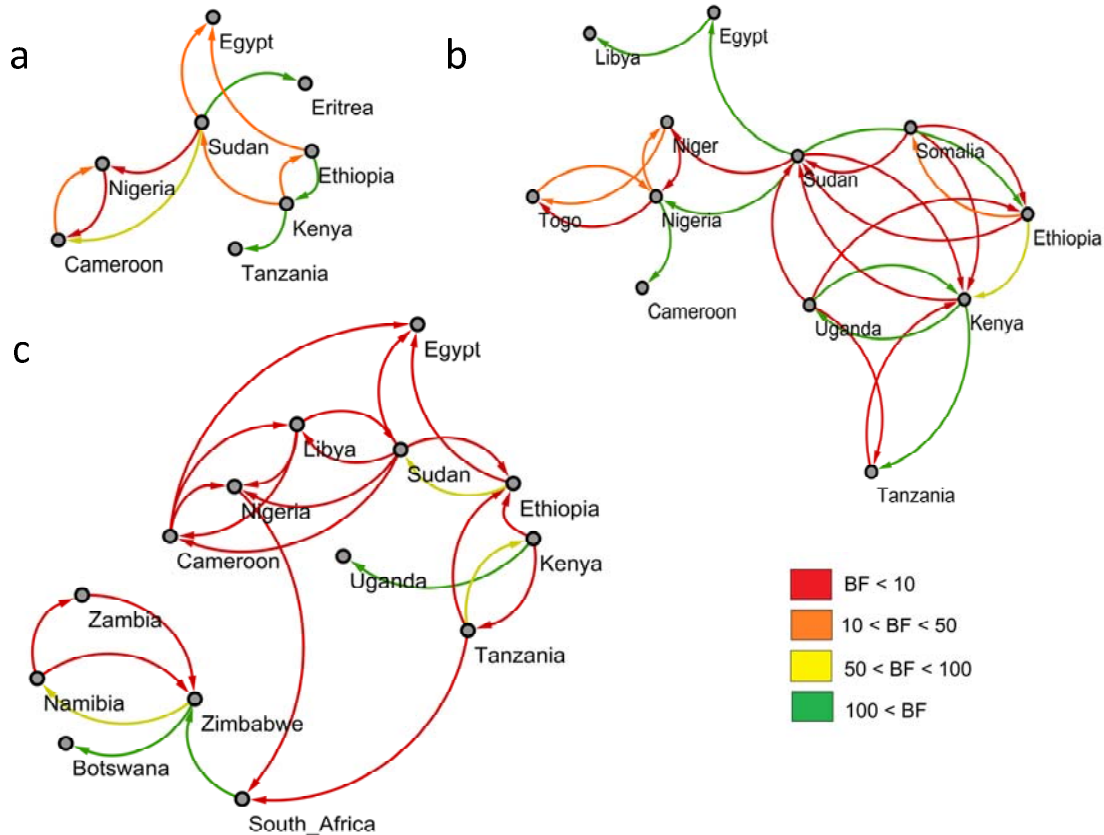
125 Five geographically defined main clades, with location posterior probabilities above 45% can be
126 observed. The first clade is exclusively composed of sequences from Botswana, Namibia and
127 Zimbabwe and seems to have its origin in the first half of the 19th century. The second clade is
128 composed of Ethiopian, Kenyan, Ugandan and Tanzanian sequences and seems to originate at the
129 transition between the 19th and 20th century. The third clade seems to have emerged at the end of the

130 18th century and is composed of Zimbabwean and all the South-African sequences. The fourth clade
131 has its TMRCA in the first half of the 19th century and is composed of sequences from Botswana,
132 Namibia and Zambia. The last clade is the most more diverse and emerged over the last century and
133 is composed of Eastern, Western and Northern African sequences (Cameroon, Egypt, Ethiopia,
134 Libya, Nigeria and Sudan). Compared to the other serotypes, the SAT2 serotype seems to have
135 appeared first in Southern Africa before moving to other parts of the continent; with a separation
136 between Southern countries and the other African countries. Whilst the presence of the virus in
137 Southern Africa seems to be geographically defined, the virus seems to circulate more freely amongst
138 the other African countries.

139 Bayesian stochastic search variable selection analysis

140

141 Using a Bayesian stochastic search variable selection (BSSVS) analysis we identified well supported
142 rates of transition between the sampled countries. The support for the rates was quantified with Bayes
143 factors (BF), and rates with $BF \geq 3$ are shown in Figure 2. Globally, the results for the serotypes A
144 and O look quite similar, with Sudan acting as a link between the Eastern, Northern and Western part
145 of the continent (see fig. 2a and 2b). For both serotypes there is a clear transmission route starting
146 from Ethiopia, passing through Kenya to Tanzania. Although most of the observed transmission
147 routes have low BF values, the situation for the SAT2 serotype is fairly similar to what is observed for
148 the serotypes A and O (see fig.2c). For SAT2 multiple transitions rates can be observed within
149 Eastern and western Africa with Sudan acting as link between the two sides. However, with only two
150 rates linking South-Africa to the rest of the continent, Southern-African countries are quite isolated
151 from the other African countries (see supp table 1,2 and 3).



152

153 **Figure 2 : Outputs of the BSSVS analysis for the three studied FMDV serotypes showing the best**
 154 **supported rates of transition between the sampled countries. The edges colours represent the relative**
 155 **strength by which the rates are supported (red: X-Y, yellow: X-Y, green: >= Z). a. For FMDV serotype A. b.**
 156 **For FMDV serotype O. c. For FMDV serotype SAT2.**

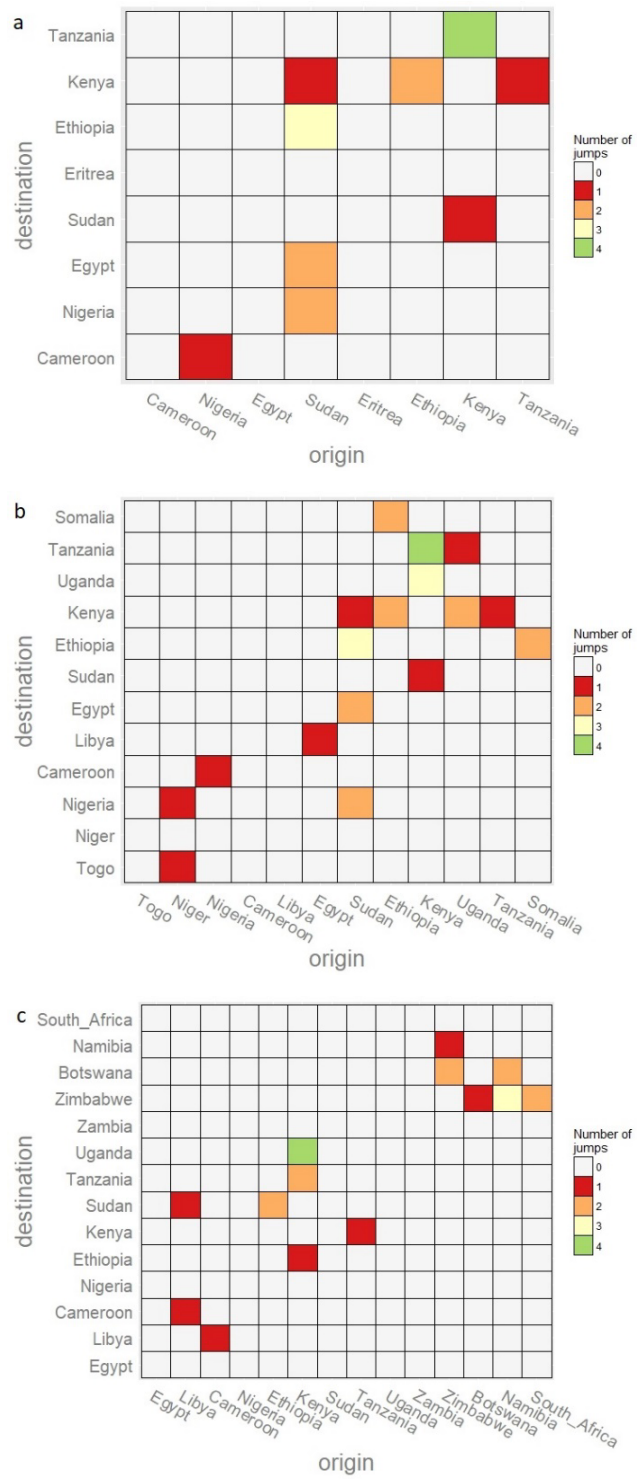
157

158 Markov jumps analysis

159

160 To complement the BSSVS analysis, an estimation of the number of transmission events between the
 161 different locations using a Markov jump analysis was performed. For both FMDV serotypes A and O
 162 we observe some transmission events starting in Kenya toward other East African countries such as
 163 Tanzania, Ethiopia and Uganda. For these two serotypes we also detected jumps from Sudan in the
 164 direction of North-Eastern and Western African countries such as Egypt, Eritrea, Cameroon and
 165 Nigeria. Therefore, it seems that for these two serotypes Kenya and Sudan act as sources for the
 166 virus, but toward different directions (see supp. table 4 and 5). For the serotype SAT2 the situation is
 167 quite different. In this case we observed fewer transition events between the different sampled

168 countries (see supp. table 6), with most of the observed transitions occurring within Eastern African
 169 and Southern Africa with no clear link between them.



170

171 **Figure 3: Heatmap showing the number of transitions between the sampled countries for the three**
 172 **studies FMDV serotypes. The heatmaps are coloured according to the number of observed transitions**
 173 **between countries. a. FMDV serotype A. b. FMDV serotype O. c. FMDV serotype SAT2**

174 **Environmental and anthropological factors affecting FMDV**
175 **diffusion**

176 Using the output from previous discrete phylogeographical analysis we isolated a FMDV serotype O
177 monophyletic clade with a time to the most recent common ancestor (TMRCA) below 25 years and a
178 high posterior probability on the location for all its nodes. The selected 46 sequences originated from
179 Kenya, Tanzania and Uganda (See supp. figure 1 and supp. table 8). Using a general linear model
180 (GLM) for the discrete location approach and the recently developed *SERAPHIM* package³⁴ for the
181 continuous location approach, we tested the impact of 13 different anthropological and environmental
182 factors (predictors) on the FMDV diffusion in Eastern Africa.

183 **Table 1 : Environmental and anthropological predictors tested for an effect on the FMDV serotype O**
184 **diffusion in Eastern Africa**

Environmental predictors	Anthropological predictor
Distance	Accessibility
Elevation	Cattle density
Precipitation	Presence of crop
Temperature	Presence of fragmented crop
Presence of forest	Human density
Presence of herbaceous vegetation	Logarithm of cattle density
	Logarithm of human density

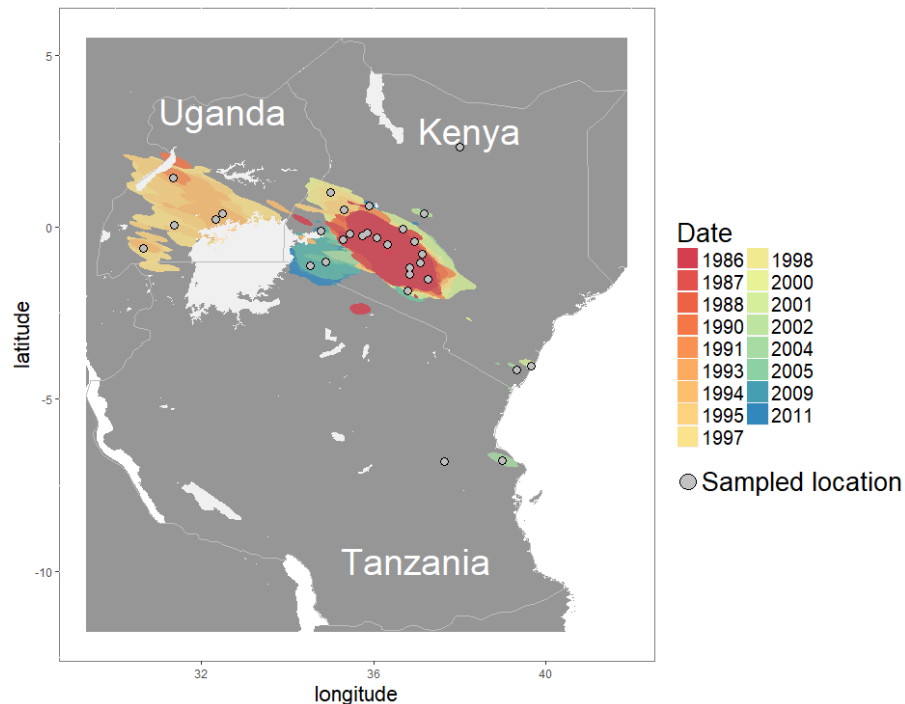
185

186 **Predictive factors for FMDV diffusion using a discrete location approach**
187

188 A generalized linear model (GLM) was used to parametrize the transition rate matrices between
189 the sampled locations as a function of our selected predictors³² on a posterior set of time-resolved
190 trees. We considered the set of predictors to be ‘conductors’ – i.e. enhancing viral diffusion, or
191 ‘resistors’ – i.e. impeding viral diffusion. We observed that the diffusion process was enhanced by
192 the average daily temperature (BF 4), the logarithm of the cattle density (BF 4) and human densities
193 (BF 9) (see table 2). It was impeded by the accessibility (BF 8), the distance between sampled
194 locations (BF 8), average amount of precipitation (BF 7) per year and by the average daily
195 temperature (BF 7) (for all the results see supp. Table 9 and 10). To gain a better understanding of
196 the impact of the average temperature and precipitation on the viral diffusion we selected different
197 thresholds of precipitation and temperature to parametrize our GLM analysis (see supp. table 11 and
198 12). We detected that low precipitation values (< 80 mL/year) were associated with an impeding

199 (negative) impact on the viral diffusion processes whereas high precipitation was associated with an
200 enhancing (positive) effect on the diffusion process. We also observed that in the case of low
201 temperature (below 22°C) a positive effect on the diffusion could be observed whereas temperatures
202 around 22°C had a negative effect on virus diffusion. Temperatures above 24°C seemed again to have
203 a positive impact on the virus spread. It was difficult to distinguish between the effects of accessibility
204 and human density because the two were strongly negatively correlated confounding the analysis
205 (see sup. table 13).

206 Predictive factors for FMDV diffusion using a continuous diffusion approach
207



208
209 **Figure 4: Map showing the continuous diffusion of the isolated clade of FMDV serotype O, with the**
210 **sampled locations as grey circles. The virus movements were reconstructed using a random walk model**
211 **with a underlying lognormal distribution**

212 Using a random walk model, we were able to reconstruct the virus diffusion in a continuous setting for
213 the isolated FMDV serotype O. Using the R package *SERAPHIM*³⁴ we evaluated the impact of the
214 predictors on the virus diffusion and observed an impeding (negative) influence of the cattle density
215 (BF 3), the presence of cropland (fragmented cropland and pure cropland areas combined) (BF4) and
216 by the presence fragmented cropland (BF6). We were not able to detect a predictor with an
217 enhancing (positive) influence on the diffusion process (see table 2).

218 To gain a better understanding of the role of the fragmented crop and cattle density we isolated the
 219 areas newly covered over the course of the infection and observed how the presence of the two
 220 predictors evolved. Overall, we noticed an opposite trend in how their densities evolved with the
 221 elapsed time (see supp.fig.2). For the fragmented crop density, high values of crop densities became
 222 more common over the course of the epidemic, with the disease moving from areas with low crop
 223 densities to areas with high crop densities. For the cattle density the opposite trend was observed
 224 with high values of cattle density more common at the start of the epidemic, the disease starting in an
 225 area where cattle densities were high and moving toward areas with lower densities of cattle. To
 226 better understand the effect that the cattle density had on virus diffusion we looked at selected areas
 227 above different thresholds of cattle density and used them as inputs in in the *SERAPHIM* package. By
 228 doing so we were able to observe that densities of cattle above 125 cattle per square km have the
 229 biggest negative impact on the virus diffusion (see supp. table 14).

230 **Table 2: Bayes factor values associated with the effect of each predictor on the connectivity between the**
 231 **sampled locations using a discrete or continuous location approach. Each predictor raster was used as**
 232 **conductance or resistance to evaluate if the predictor have a positive or negative effect on the viral**
 233 **genetic connectivity.**

Predictor	<i>Discrete locations</i>		<i>Continuous locations</i>	
	Conductance	Resistance	Conductance	Resistance
	Bayes factor	Bayes factor	Bayes factor	Bayes factor
Accessibility	0	8	0	0
Cattle density difference	-	0	-	-
Cattle density	2	0	0	3
Presence of crop	0	0	1	0
Presence of crop (combined)	0	0	0	4
Distance	-	8	-	-
Elevation	1	0	0	1
Presence of forest	1	0	1	0
Presence of fragmented crop	1	0	0	6
Human density	1	0	1	0
Precipitation	2	7	0	0
Presence of herbaceous vegetation	1	0	0	0
Temperature	4	7	0	1
Logarithm Cattle density	4	1	1	0
Logarithm Human density	9	2	0	0

234

235 **DISCUSSION**

236 In this paper we have applied the most recent existing phylogenetic methods on all available African
237 FMDV VP1 sequences for the serotypes A, O and SAT2. Our work has some limitations, especially
238 regarding the limited availability of sequences. Our sampling is obviously unbalanced as it is based on
239 submissions by individual countries or ad hoc research projects and the effect that it has on the
240 results quality is uncertain. Increasing the number of available FMDV sequences from diverse
241 locations and hosts would help to develop models that better represent the diffusion of FMDV in Africa
242 and lead to better environmental and anthropological effect estimation.

243 The estimated substitution rates of 4.67×10^{-3} , 3.69×10^{-3} and 1.1×10^{-3} mutations per site per year for
244 the serotypes A, O and SAT2 from our results are similar to previous estimates of 4.26×10^{-3} ,
245 3.14×10^{-3} and 1.07×10^{-3} mutations per site per year for the same serotypes as found by Tully et al³⁵.
246 Overall, we observed similar evolutionary patterns for both FMDV serotypes A and O. Those
247 serotypes seem to have appeared in Eastern Africa around 1930. Our results pointed to the possible
248 role of Kenya as a viral source for East African countries and the role of Sudan as a link between East
249 Africa and North-East Africa. The evolution of the SAT2 serotype seems to be quite different.
250 According to our results, this serotype seems to have emerged around 1583. Over the 5 main clades
251 that compose the SAT2 phylogeny only two emerged outside Southern Africa (clades 2 and 5). Those
252 two clades involved Eastern (Kenya, Ethiopia and Tanzania), or Western/Central/Northern African
253 countries (Libya, Nigeria, Cameroon and Egypt) with Sudan being the only country present in both
254 clades. Additionally, these two clades have emerged more recently than those involving Southern
255 African countries (second part of 19th century/early 20th century for the clade 2 and 5 against late 18th
256 century/early 19th century for the other clades).

257 For the FMDV serotypes A and O the observation of well supported rates of viral transmission
258 between Eastern Africa and Western Africa can be explained by the existence of commercial routes
259 between those areas. It is indeed acknowledged that livestock trades play an important role in FMDV
260 circulation in sub-Saharan Africa, through trading routes exist between the Horn of Africa and Eastern
261 Africa, with Sudan acting as key commercial intermediate⁸. Additionally, the existence of a relatively
262 recent common ancestor for the FMDV serotypes A and O is further support for the idea that these
263 serotypes were imported into Africa at the start of the 19th century through livestock trade from Asia
264 and Europe¹¹.

265 The SAT2 serotype analysis shows the signs of the impact that the African rinderpest epidemics had
266 on FMDV circulation in Africa. Although FMDV was first reported in southern Africa in 1795 it likely
267 had coevolved with buffalo over millennia resulting in a large diverse viral pool, but the rinderpest
268 epidemic decimated almost all FMDV potential carriers and probably pushed it through a huge
269 bottleneck³⁶. It is thought that FMDV re-emerged from wild buffalo population that survived the
270 rinderpest epidemic, before being reported again in 1932 in Southern Africa¹¹. Consistent with this
271 hypothesis, of all the clades present in the reconstructed phylogeny, only those originating from
272 Southern African countries have an TMRCA older than the African rinderpest epidemic (clades 1, 3
273 and 4). The SAT2 serotype probably spread into non-Southern African countries through infected
274 livestock movements, increasing the virus mobility and explaining the more recent TMRCA and the
275 highest country diversity observed in the more recent clades (clades 2 and 5). Additionally, with the
276 observation of similar transition rates patterns amongst the three serotypes, the BSSVS analysis
277 suggest that over the last hundred years the SAT2 circulation between African countries was mainly
278 driven by domestic animal movements, with the relative isolation of southern Africa being the result of
279 the different control measures in place in this region³⁷.

280 Using both a discrete and continuous framework, we looked at the effect that diverse environmental
281 and anthropological factors had on the diffusion of an isolated FMDV serotype O clade that circulated
282 in Kenya, Uganda and Tanzania. The results of the discrete approaches suggest that the FMDV
283 diffusion is facilitated by low average daily temperature (<22°C), high averages precipitations (>80
284 mL/year) as well as high human and cattle densities. We were also able to observe that the virus
285 diffusion was negatively impacted by the accessibility (long travel time needed to join the closest
286 major city) in addition to high daily temperatures and low average precipitations. Since lower
287 temperatures and higher humidity values are usually associated with a longer virus survivability in the
288 environment³⁸, our results may suggest a more important role than what was previously believed of
289 the indirect transmission through viral persistence in the environment for FMV in this region.
290 Additionally, with the viral diffusion being positively affected by high cattle and human densities and
291 negatively affected by large accessibility values, anthropological activities seem to have an impact on
292 the virus diffusion. These observations could be the consequence of infected herds of cattle moving
293 from smaller rural localities toward nearby larger cities with cattle markets^{14,39}.

294 Regarding the effect of the different selected predictors on virus diffusion in a continuous setting, our
295 results suggest that cattle densities above 125 cattle per km² and the presence of cropland (pure
296 cropland or mixed with other types of land) both have a negative impact on virus diffusion. Ours
297 results suggest that the virus had difficulties to spread beyond the geographic region located at the
298 root of the tree, where high cattle densities and low crop densities were present and to spread to
299 areas with low cattle densities but high crop densities, presumably due to lack of suitable hosts.
300 However, it is difficult to know exactly whether it is the cattle or crop density that had the most impact
301 due to high correlation of the two variables at the time and in the region of origin.

302 The location uncertainty found at the root of the continuous tree could explain the differences between
303 the discrete and continuous methods in estimating the effect of the cattle density on virus diffusion.
304 For our analysis, this uncertainty seems to be translated by the *SEREPHIM* programme as a period
305 where the virus is almost not moving. This uncertainty seems to drive *SEREPHIM* to the conclusion
306 that the high cattle densities found near the origin of the epidemic are related to this lack of movement
307 and therefore estimate that they have a negative influence on the virus diffusion. Although we suspect
308 a link between the cattle density and the location of emergence of the analysed clade, we think that
309 the continuous analysis does not offer the resolution needed to understand that relation (i.e. the
310 spatial HPD confidence interval is too large). By parameterizing each rate of among-location
311 movement as a function of predictors, the discrete approach seems therefore more appropriate to
312 characterise the environmental and anthropological effect of the virus diffusion in this endemic
313 situation.

314 In conclusion, the reconstituted phylogeographical tree pattern for the FMDV serotypes A, O and
315 SAT2 reflects a situation where the recent FMDV circulation is mainly driven by commercial
316 exchanges, through pastoral herd movements, and where wildlife seems to have almost no influence
317 on the intra-continental circulation of the disease. However, the observed patterns for SAT2 reflects a
318 situation where wildlife (wild buffaloes) constitute the original host of the serotype, whilst the
319 observations for A and O suggest that those serotypes were imported in Africa at the start of the 19th
320 century. We observed that indirect transmission through the environment and direct transmission
321 through anthropological activities had an enhancing effect on the virus diffusion in Eastern Africa.

322 **MATERIALS AND METHODS**

323 **Data Collection**

324 To obtain a comprehensive genetic dataset we first retrieved all available African FMDV A, FMDV O
325 and FMDV SAT2 genetic sequences in Genbank (accessed on the 15/12/2016). From these datasets
326 we selected all VP1 sequences with at least information on the country of sampling and the year of
327 sampling. In total we gathered 191 FMDV A, 351 FMDV O and 477 FMDV SAT2 sequences. The
328 sequences were aligned using Multiple Alignment Fast Fourier transformation (*MAFFT*)⁴⁰. Potential
329 recombinant sequences were detected with RDP4 software and any such sequences were
330 removed⁴¹.

331 To reduce the effect of potential sampling bias we removed all sequences coming from countries with
332 less than three sequences available (in the whole-time span) and ran a stratified subsampling
333 procedure to allow a maximum of three sequences per country of origin and per month. The spatial
334 coordinates of sampling for each sequence was retrieved using the *GGMAP* package in R and the
335 most precise sampling localisation name available for each sequence⁴². The final FMDV A dataset
336 was composed of 107 sequences from eight countries, dates ranging from 1966 to 2016. The final
337 FMDV O dataset was composed of 192 sequences from 12 countries, dates ranging from 1964 to
338 2016. The final FMD SAT2 dataset was composed of 135 sequences from 15 countries, dates
339 ranging from 1970 to 2015 (see supp tables 15,16 and 17).

340 **Bayesian Evolutionary Inference**

341 Discrete phylogeographic tree inference

342 Time-scaled phylogenetic trees were inferred using *BEAST* 1.8 with the *BEAGLE* library⁴³, and
343 different substitution clock and population evolution models were evaluated by estimating their
344 marginal likelihoods using the Akaike's Information Criterion for MCMC samples (AICM) in Tracer 1.6.
345 Ultimately a general-time-reversible (GTR) model with site to site rate variation between two
346 categories was selected as nucleotide substitution model⁴⁴ with a Bayesian skygrid population model
347 and a relaxed uncorrelated log-normal molecular clock model were chosen to model the evolution of
348 the FMDA and FMDO serotypes. For the FMD SAT2 serotype a HKY nucleotide substitution model
349 with a constant clock model with a Bayesian skygrid population model were chosen to model its

350 evolution^{45,46}. Posterior sets of trees were generated for each serotype by combining at least 2
351 independent Markov Chain Monte Carlo run of 40 million steps sampling every thousand with 10%
352 burn-in.

353 We first reconstructed the time-scaled phylogenetic trees for the three studied serotypes. Thereafter
354 to reduce the computation time of the GLM and the spatial diffusion analyses we estimated the spatial
355 model components using subsets of 1000 trees from the original posterior distributions of trees as
356 input empirical tree distributions. We used *TreeAnnotator* to summarize maximum clade credibility
357 (MCC) trees and *FigTree* version 1.4.1 to visualize the annotated trees^{47,48}. The software *SPREAD3*
358 and *Cytoscape* were used to identify and visualize the well supported rates of transmission through a
359 Bayes factor test⁴⁹.

360 For the three serotypes we reconstituted the discrete transition events between the different sampled
361 countries through the whole phylogeny using the “migration model”. Therefore, an asymmetric
362 continuous-time Markov chain (CTMC) model with an incorporated Bayesian stochastic search
363 variable selection (BSSVS) was used to determine which set of transition rates sufficiently
364 summarizes the epidemiological connectivity between the countries³⁰. A posterior inference of the
365 complete Markov jump history through the whole genealogy was also performed, in order to quantify
366 state transitions and the time spent in each country by the virus.

367 Environmental and anthropological effect estimation

368 *Monophyletic clade selection*

369 Using the previously reconstructed discrete phylogeographic tree of the FMDV O serotype we
370 selected a monophyletic clade with a MRCA under 25 years and a posterior probability over 50 % on
371 the location for all its nodes. To avoid uncertainty in the predictor effect estimation analysis we
372 removed all sequences connected to branches with length more than 10 years. At the end of the
373 process the dataset was composed of 46 FMDV O sequences coming from 31 locations across
374 Kenya, Uganda and Tanzania.

375 *Generation of predictive factors of FMDV diffusion*

376 A Generalised Linear Model (GLM) extension of the discrete approach was used to test and quantify
377 the enhancing (positive) or impeding (negative) effect of potential predictors on the viral diffusion
378 process³². This model parametrizes the transmission rate matrix between discrete locations as a log

379 linear function of the potential predictive factor matrices. While reconstructing the phylogeographic
380 history the model performs Bayesian model averaging to determine which combination of predictor
381 matrices are the best to explain the spatial diffusion process. For each predictor a Bayes factor (BF)
382 value is calculated based on the ratio of posterior to prior probabilities of inclusion⁵⁰.

383 The different predictors of FMD diffusion considered were: the accessibility to the sampled location,
384 cattle density, crop density, the elevation of the location, the forest density, the human density, the
385 average yearly precipitation, the shrubland area density, the average daily temperature (for the
386 provenance see sup, table 7). Each potential predictor was retrieved as a raster matrix, representing
387 the predictor spatial localisation, and aggregated to a resolution of 0.08 by 0.08, corresponding to
388 pixels of approximately 8 km by 8 km.

389 The circuitscape software was used to determine the predictors values used in our GLM analysis⁵¹.
390 For each predictor, two predictor values were generated, one using the raster as resistance values
391 (impeding the viral diffusion) and the other using the raster as conductance value (enhancing the viral
392 diffusion). To obtain those values we used a circuit theory approach to estimate modified distances,
393 used as predictor values, between each pair of locations using the raster values as heterogeneity
394 factors⁵⁰. Consequently, if a raster was used as a resistance surface, we would estimate large
395 predictor values between the locations separated by high raster values and small predictor values
396 between the locations separated by small raster values. Prior to their inclusion in the GLM analyses
397 the predictor values were log transformed and standardised. Each analysis was run by comparing the
398 effect of a predictor with a null predictor, corresponding to a random raster.

399 Complementary to the discrete GLM approach, we tested and quantified the effect of the potential
400 predictors using a continuous coordinate approach. Therefore, we inferred the diffusion of the virus
401 using a random walk model of diffusion and used the *SERAPHIM* package to test and estimate the
402 effect of the predictors on the virus diffusion^{31,34}. Like the discrete approach, *SERAPHIM* estimates a
403 modified distance for each pair of locations found at the start and end of the phylogeny branches. The
404 correlation between the time spent on each branch and the estimated distance value is then
405 estimated. The statistical significance of this correlation is tested using a randomized phylogeny and
406 expressed in the form of a BF³⁴.

407 REFERENCES

- 408 1. Alexandersen, S., Zhang, Z., Donaldson, A. I. & Garland, A. J. M. The pathogenesis and diagnosis of foot-and-mouth
409 disease. *J. Comp. Pathol.* **129**, 1–36 (2003).
- 410 2. Belsham, G. J. DISTINCTIVE FEATURES OF FOOT-AND-MOUTH DISEASE VIRUS, A MEMBER OF THE
411 PICORNAVIRUS FAMILY; ASPECTS OF VIRUS PROTEIN SYNTHESIS, PROTEIN PROCESSING AND
412 STRUCTURE. *Prog. Biophys. molec. Biol.*, **60**, 241–260 (1993).
- 413 3. Fuquay, J. W., Schrijver, R. S. & Vosloo, W. Infectious Diseases: Foot-and-Mouth Disease. *Encycl. Dairy Sci.* 160–
414 167 (2011). doi:<http://dx.doi.org/10.1016/B978-0-12-374407-4.00132-1>
- 415 4. Domingo, E., Baranowski, E., Escarmís, C. & Sobrino, F. Foot-and-mouth disease virus. *Comp. Immunol. Microbiol.*
416 *Infect. Dis.* **25**, 297–308 (2002).
- 417 5. Alexandersen, S. & Mowat, N. Foot-and-mouth disease: host range and pathogenesis. *Curr. Top. Microbiol. Immunol.*
418 **288**, 9–42 (2005).
- 419 6. Genevieve V. Weaver, Joseph Domenech, Alex R. Thiermann, and W. B. K. FOOT AND MOUTH DISEASE A LOOK
420 FROM THE WILD SIDE. *J. Wildl. Dis.* **49**, 759–785 (2013).
- 421 7. Donaldson, A. I., Alexandersen, S., Sorensen, J. H. & Mikkelsen, T. Relative risks of the uncontrollable (airborne)
422 spread of FMD by different species. *Vet. Rec.* **148**, 602–604 (2001).
- 423 8. Di Nardo, a, Knowles, N. J. & Paton, D. J. Combining livestock trade patterns with phylogenetics to help understand
424 the spread of foot and mouth disease in sub-Saharan Africa, the East and Southeast Asia. *Middle. Rev. Sci. Tech.* **30**,
425 63–85 (2011).
- 426 9. Robinson, T. P. *et al.* *Global livestock production systems.* (2011).
- 427 10. Lyons, N. A. *et al.* Impact of foot-and-mouth disease on milk production on a large-scale dairy farm in Kenya. *Prev.*
428 *Vet. Med.* **120**, 177–186 (2015).
- 429 11. Casey, M. B. *et al.* *Patterns of Foot-and-Mouth Disease Virus Distribution in Africa: The Role of Livestock and Wildlife*
430 *in Virus Emergence. The Role of Animals in Emerging Viral Diseases* (Elsevier, 2013). doi:10.1016/B978-0-12-
431 405191-1.00002-8
- 432 12. Vosloo, W., Bastos, A. D., Sangare, O., Hargreaves, S. K. & Thomson, G. R. Review of the status and control of foot
433 and mouth disease in sub-Saharan Africa. *Rev Sci Tech* **21**, 437–449 (2002).
- 434 13. Tekleghiorghis, T., Moormann, R. J. M., Weerdmeester, K. & Dekker, A. Foot-and-mouth Disease Transmission in
435 Africa: Implications for Control, a Review. *Transbound. Emerg. Dis.* **63**, 136–151 (2016).
- 436 14. Motta, P. *et al.* Implications of the cattle trade network in Cameroon for regional disease prevention and control.
437 (2017). doi:10.1038/srep43932
- 438 15. Vosloo, W., Bastos, A. D. S., Sangare, O., Hargreaves, S. K. & Thomson, G. R. Review of the status and control of

- 439 foot and mouth disease in sub-Saharan Africa. *Rev. Sci. Tech.* **21**, 437–49 (2002).
- 440 16. Miguel, E. *et al.* Contacts and foot and mouth disease transmission from wild to domestic bovines in Africa. *Ecosphere*
441 **4**, 1–32 (2013).
- 442 17. Bastos, A. D., Boshoff, C. I., Keet, D. F., Bengis, R. G. & Thomson, G. R. Natural transmission of foot-and-mouth
443 disease virus between African buffalo (*Syncerus caffer*) and impala (*Aepyceros melampus*) in the Kruger National
444 Park, South Africa. *Epidemiol. Infect.* **124**, 591–8 (2000).
- 445 18. Thomson, G. R., Vosloo, W. & Bastos, A. D. S. Foot and mouth disease in wildlife. *Virus Res.* **91**, 145–161 (2003).
- 446 19. Casey-Bryars, M. *et al.* Waves of endemic foot-and-mouth disease in eastern Africa suggest feasibility of proactive
447 vaccination approaches. *Nat. Ecol. Evol.* **1** (2018). doi:10.1038/s41559-018-0636-x
- 448 20. Allepuz, A. *et al.* Risk factors for foot-and-mouth disease in Tanzania, 2001-2006. *Transbound. Emerg. Dis.* **62**, 127–
449 136 (2015).
- 450 21. Wungak, Y. S., Olugasa, B. O., Ishola, O. O., Lazarus, D. D. and Ularamu, G. H. Foot-and-mouth disease (FMD)
451 prevalence and exposure factors associated with seropositivity of cattle in north-central , Nigeria. **15**, 1224–1232
452 (2016).
- 453 22. Dean, A. S. *et al.* Potential Risk of Regional Disease Spread in West Africa through Cross-Border Cattle Trade. *PLoS*
454 *One* **8**, 1–9 (2013).
- 455 23. Hamoonga, R., Stevenson, M. A., Allepuz, A., Carpenter, T. E. & Sinkala, Y. Risk factors for foot-and-mouth disease
456 in Zambia, 1981-2012. *Prev. Vet. Med.* **114**, 64–71 (2014).
- 457 24. Bessell, P. R., Shaw, D. J., Savill, N. J. & Woolhouse, M. E. Geographic and topographic determinants of local FMD
458 transmission applied to the 2001 UK FMD epidemic. *BMC Vet. Res.* **4**, 40 (2008).
- 459 25. Dion, E., VanSchalkwyk, L. & Lambin, E. F. The landscape epidemiology of foot-and-mouth disease in South Africa: A
460 spatially explicit multi-agent simulation. *Ecol. Modell.* **222**, 2059–2072 (2011).
- 461 26. Flood, J. S., Porphyre, T., Tildesley, M. J. & Woolhouse, M. E. The performance of approximations of farm contiguity
462 compared to contiguity defined using detailed geographical information in two sample areas in Scotland: implications
463 for foot-and-mouth disease modelling. *BMC Vet. Res.* **9**, 198 (2013).
- 464 27. Kühnert, D., Wu, C.-H. & Drummond, A. J. Phylogenetic and epidemic modeling of rapidly evolving infectious
465 diseases. *Infect. Genet. Evol.* **11**, 1825–1841 (2011).
- 466 28. Drummond, A., Pybus, O. G. & Rambaut, A. Inference of Viral Evolutionary Rates from Molecular Sequences. *Adv.*
467 *Parasitol.* **54**, 331–358 (2003).
- 468 29. Pybus, O. G. & Rambaut, A. Evolutionary analysis of the dynamics of viral infectious disease. *Nat. Rev. Genet.* **10**,
469 540–550 (2009).

- 470 30. Lemey, P., Rambaut, A., Drummond, A. J. & Suchard, M. A. Bayesian Phylogeography Finds Its Roots. *PLoS*
471 *Comput. Biol.* **5**, e1000520 (2009).
- 472 31. Lemey, P., Rambaut, A., Welch, J. J. & Suchard, M. A. Phylogeography Takes a Relaxed Random Walk in
473 Continuous Space and Time. *Mol. Biol. Evol.* **27**, 1877–1885 (2010).
- 474 32. Lemey, P. *et al.* Unifying Viral Genetics and Human Transportation Data to Predict the Global Transmission Dynamics
475 of Human Influenza H3N2. *PLoS Pathog.* **10**, e1003932 (2014).
- 476 33. Dellicour, S., Rose, R. & Pybus, O. G. Explaining the geographic spread of emerging viruses: a new framework for
477 comparing viral genetic information and environmental landscape data. *BMC Bioinformatics in press*, 1–12 (2016).
- 478 34. Dellicour, S., Rose, R., Faria, N. R., Lemey, P. & Pybus, O. G. SERAPHIM: studying environmental rasters and
479 phylogenetically-informed movements. *Bioinformatics* btw384 (2016). doi:10.1093/BIOINFORMATICS/BTW384
- 480 35. Tully, D. C. & Fares, M. A. The tale of a modern animal plague: Tracing the evolutionary history and determining the
481 time-scale for foot and mouth disease virus. *Virology* **382**, 250–256 (2008).
- 482 36. Knowles, N. J. FMD history. (1990). Available at: http://www.picornaviridae.com/aphthovirus/fmdv/fmd_history.htm.
483 (Accessed: 25th May 2018)
- 484 37. Perry, B. D. *et al.* The impact and poverty reduction implications of foot and mouth disease control in southern Africa
485 with special reference to Zimbabwe. (2003).
- 486 38. Bartley, L. M., Donnelly, C. A. & Anderson, R. M. Review of foot-and-mouth disease virus survival in animal excretions
487 and on fomites. *Vet. Rec.* **151**, 667–9 (2002).
- 488 39. Robinson, S. E. & Christley, R. M. Exploring the role of auction markets in cattle movements within Great Britain. *Prev.*
489 *Vet. Med.* **81**, 21–37 (2007).
- 490 40. Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: multiple sequence alignment, interactive sequence
491 choice and visualization. *Brief. Bioinform.* (2017). doi:10.1093/bib/bbx108
- 492 41. Martin, D. P., Murrell, B., Golden, M., Khoosal, A. & Muhire, B. RDP4: Detection and analysis of recombination
493 patterns in virus genomes. *Virus Evol.* **1**, vev003 (2015).
- 494 42. Kahle, D. & Wickham, H. ggmap: Spatial Visualization with ggplot2.
- 495 43. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214
496 (2007).
- 497 44. Tavare, S. Some probabilistic and statistical problems in the analysis of dna sequences. *Lect. Math. Life Sci.*
498 (*American Math. Soc.* **17**, 57–86 (1986).
- 499 45. Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. Relaxed Phylogenetics and Dating with Confidence.
500 *PLoS Biol.* **4**, e88 (2006).

- 501 46. Gill, M. S. *et al.* Improving bayesian population dynamics inference: A coalescent-based model for multiple loci. *Mol.*
502 *Biol. Evol.* **30**, 713–724 (2013).
- 503 47. Rambaut, A. FigTree. (2009). Available at: <http://tree.bio.ed.ac.uk/software/figtree/>. (Accessed: 25th April 2018)
- 504 48. Bouckaert, R. *et al.* BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Comput. Biol.* **10**,
505 e1003537 (2014).
- 506 49. Bielejec, F. *et al.* Spread3: Interactive Visualization of Spatiotemporal History and Trait Evolutionary Processes. *Mol.*
507 *Biol. Evol.* **33**, 2167–2169 (2016).
- 508 50. Trovão, N. S. *et al.* Host ecology determines the dispersal patterns of a plant virus. *Virus Evol.* **1**, vev016 (2015).
- 509 51. McRae, B. H. Isolation By Resistance. *Evolution (N. Y.)* **60**, 1551 (2006).

510

511 Acknowledgements

512 The study was supported by Roslin Institute Strategic grant funding from the the Biotechnology and
513 Biological Sciences Research Council, UK (BBS/E/D/20002173). SL is supported by a Chancellor's
514 Fellowship from the University of Edinburgh, and FD is supported by a Chancellor's Fellowship
515 studentship.

516

517 Author contributions

518 FD, SL and MB conceived the study; FD analysed and interpreted the data, and drafted the
519 manuscript; SL and MB provided guidance on the study design and the analysis, and contributed to
520 interpreting the data and scientific discussions. All authors read and approved the final manuscript.

521

522 Competing interests

523 The authors declare no competing interests.

524

525 Additional information

526 Early pre-print of this manuscript on BioRxiv doi: <https://doi.org/10.1101/358044>

527 <https://www.biorxiv.org/content/early/2018/06/28/358044.1>

528

529

530

531

532

533