Circulation of Foot-and-Mouth Disease Virus in Africa and identification of the underlying constraints using Phylogeographic methods

Florian Duchatel[1*], Dr. Mark Bronsvoort[1], Dr. Samantha Lycett[1]

[1]The University of Edinburgh, Roslin Institute, Easter Bush Campus, Midlothian EH25 9RG

*Corresponding author, s1576652@ed.ac.uk

ABSTRACT

Foot and mouth disease is endemic in sub-Saharan Africa and can lead to important and continuous economic losses for the infected countries. Due the complexity of the disease epidemiology and the lack of data there is a need to use novel approaches to fill the gaps in our understanding of the circulation of FMD in this continent. Using a phylogeographical approach we reconstructed the circulation of foot and mouth disease virus serotypes A, O and SAT2 in Africa and evaluated the influence of potential environmental and anthropological predictors on the virus diffusion. Our results show that the circulation of serotypes A and O seems mainly driven by commercial exchanges while the SAT2 serotype origin can be found in wildlife populations. Both indirect transmission through persistence in the environment and anthropological activities such as cattle movements seem to have an impact on the circulation of the serotype O.

INTRODUCTION

Foot and mouth disease (FMD) affects more than 70 species of cloven-hoofed animals [1]. The disease is characterised by the development of vesicles in and around the mouth, on the feet and other sites of the skin [2]. The causal agent is a positive-sense, single-stranded RNA virus of the Picornaviridae family [3] called foot-and-mouth disease virus (FMDV). Its genome encodes the information for 4 structural proteins (VP1-4) and 8 non-structural proteins (7 proteases and one RNA polymerase). Based on the level of cross protection between strains, the virus can be divided in seven serotypes, O, A, C, Southern African Territories [SAT] 1, SAT 2, SAT 3 and Asia 1 [4,5] which are clinically indistinguishable from each other but which have different epidemiology. The hosts that are considered to play an active role in the epidemiology of the virus are cattle, buffaloes, pigs, sheep and goats [6].

FMD susceptibility varies according to the infected host and the type of strain involved. The severity of the infection depends of the amount of virus inoculated, the serotype, the host species and the individual immunity [7]. The commonest route of infection for a new host is by direct contact with an infected animal [1,8]. The infection may also occur indirectly by contacts with contaminated surfaces or products, such as infected personnel, vehicle or fomites [6]. Movement of animals and animal products are considered to play an important role in the disease circulation in endemic areas and are considered the main factors for FMDV transboundary spread [9].

FMD has been eradicated in many high income countries but is still endemic in numerous low and middle income countries (LMICs) [10] particularly in Africa and South and East Asia. Although FMDV has a low mortality rate the disease causes significant productivity losses that may lead to important and continuous economic losses for the farmers and the states [9]. However, there is still a lack of data to quantify the impact the disease has on the economy of the endemically infected countries [11] though we start to have some information on its impact on large scale dairy farm LMICs [12].

FMD viruses are endemic in most of sub-Saharan Africa with an epidemiology considered to be more complex than in any other regions of the world [13]. However, very few statistics on the disease incidence and circulation are available since most outbreaks in endemic areas are unrecorded, there is a general lack of surveillance and animal movements and traceability are not well established. Further, there are few studies of animal trade and little understanding of seasonal transhumance migrations of nomadic and pastoralist herds [9,13–15]. Therefore, there is a crucial need for novel approaches to use existing data to fill the gaps in our understanding of the circulation of FMD in sub-Saharan African.

Additionally, many wildlife species can be infected by the virus in Africa [7]. Amongst all the potential wild hosts, only buffalo and impala were implicated in the transmission of FMDV to domestic cattle [16]. In South Africa the Cape buffalo is suspected to be the primary reservoir and the main source of SAT serotypes of FMD viruses. However, there are still many questions regarding the role of buffalo as a viral source for livestock outbreaks, how those populations sustain endemic cycles of infection in livestock and their role in the epidemiology of the FMDV O, A and C serotypes [11,14,17,18].

Many environmental characteristics and attributes such as the landscape, vegetation, natural barriers to animal movements such as roads, rivers or mountains have the potential to influence the dynamics and

circulation of FMD in Africa [19–21]. Previous studies have shown that different factors have been formerly linked with an increased risk of FMD transmission to domestic animals: an increase in contact between wild buffaloes and domestic cattle, the presence of humans, movements extent of the wild and domestic animals and the distance to road, railways and international borders.

As a single stranded RNA virus and due to its error prone replication cycles, FMDVs have a high mutation rate. These mutations provide information on the ecological processes and population events that shaped the virus evolution and that were not directly observed. These processes, along with other evolutionary parameters, can be modelized while reconstructing the phylogenetic trees[22,23]. Furthermore, by combining genetic data and spatial information, phylogeographical trees reconstruction can estimate the unobserved geographical circulation of a pathogen[24]. Therefore the virus movements can be modelled as discrete transmission events between the sampled location[25] or as a continuous process using different random walk diffusion models [26]. Recently both discrete and continuous approaches have been extended to test and quantify the contribution of potential environmental and anthropological parameters (predictors of viral diffusion) that might influence the spread and circulation of the studied pathogen [27,28].

The aim of this paper is to gain a better understanding of the circulation of FMDV in Africa using a detailed discrete phylogeographical analysis of serotypes A, O and SAT2 sequences. Moreover the influence of 13 potential environmental and anthropological predictors of virus diffusion were quantified and tested using both discrete and continuous approaches [27,29].

RESULTS

Discrete phylogenetic analysis

Evolutionary parameters estimation

Using a relaxed log-normal clock and a skygrid population model for the serotype A and O and a constant clock and skygrid population model for the serotype SAT2 we estimated diverse evolutionary parameters and reconstructed the phylogenetical tree for three serotypes. Overall, we observed a mean mutation rate of 4.67×10^{-3} substitutions per site per year and 3.69×10^{-3} for the serotypes A and O. We also estimated a significantly slower mutation rate of 1.1×10^{-3} for the serotype SAT2 (see supp. table 11.).

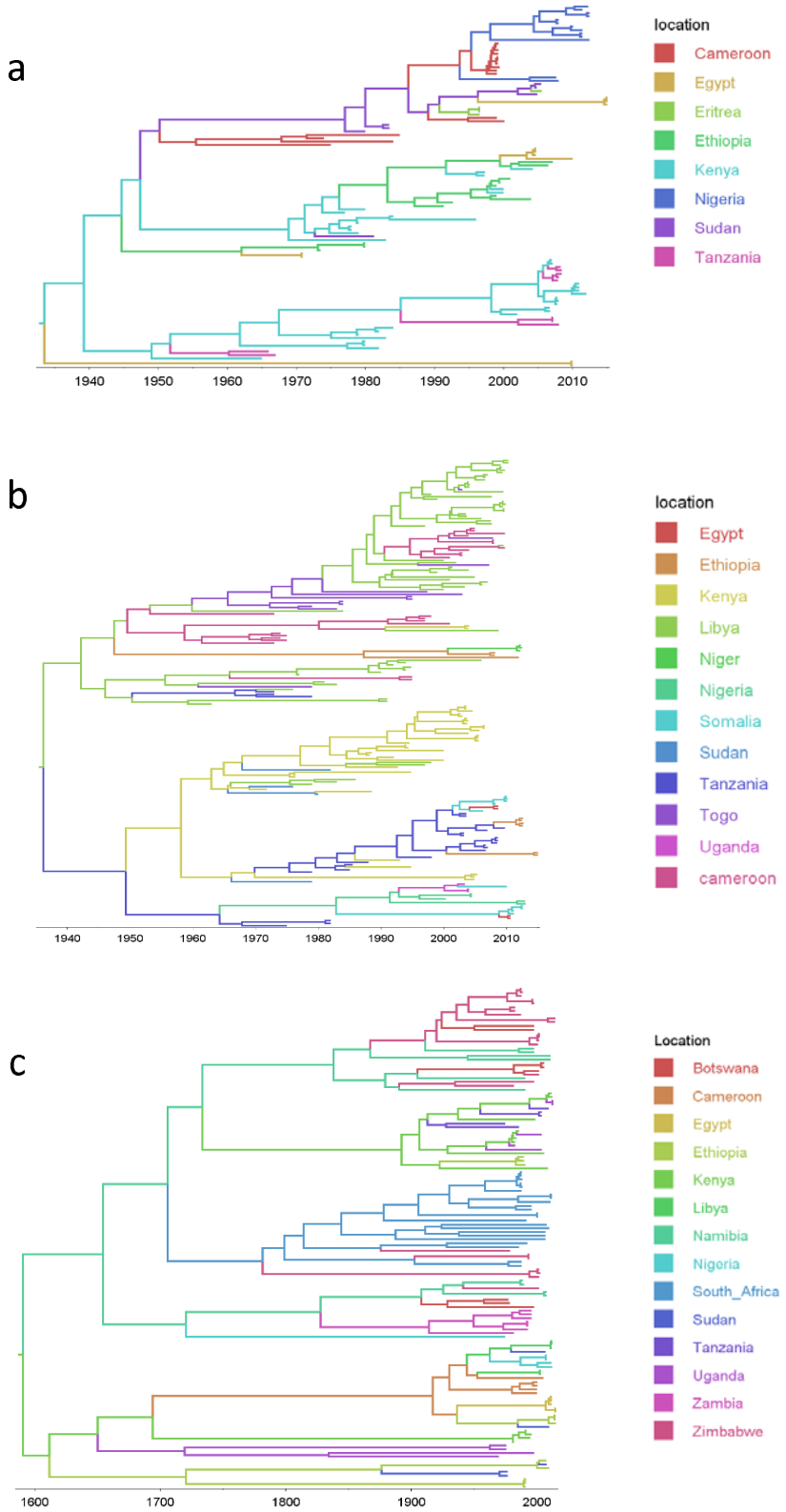

Figure 1: Bayesian MCC time scaled phylogeographic tree for the three studied serotypes. a. Bayesian phylogeographic tree for the serotype A using 107 VP1 sequences. b. Bayesian phylogeographic tree for the serotype O using 192 VP1 sequences. c. Bayesian phylogeographic tree for the serotype SAT2 using 135 VP1 sequences. The phylogeny branches are coloured according to their descendent nodes location with the key for colours shown on the right.

Phylogeographic tree reconstruction for the serotype A

According to the reconstructed phylogeographic tree the serotype A viruses probably emerged around 1926 (1889.6 – 1950 95%HPD) in the eastern part of Africa with high posteriors probabilities for Kenya (49,83 %) and Ethiopia (35,95 %) as the geographical origin of the disease (see fig. 1a). For this serotype there is no clear clade separation between the western and eastern sides of Africa with one clade combining all the western African sequences as well as sequences from Sudan, Ethiopia and Egypt. Even though few transmissions events are observable between the two sides of Africa, all of them are involving Sudan as link between them.

Phylogeographic tree reconstruction for the serotype O

The FMDV serotype O was probably first observed around 1937 (1921 – 1952 95%HPD) in the eastern part of Africa with high posteriors probabilities for Kenya (61,49 %), Sudan (17.15 %) and Uganda (11.42 %) as the geographical origin of the disease (see fig.1b). For this serotype the reconstructed phylogeographical tree is composed of four geographically defined main clades. The first clade is almost entirely composed of Kenyan, Tanzanian and Ugandans sequences with only a few transmissions to other countries. The second clade is mostly situated in Ethiopia with few transitions to Kenya and Somalia. The third clade is centred in Sudan with incursions into Nigeria, Cameroon, Egypt and Ethiopia. The last clade is centred in West and Central African countries (Cameroon, Nigeria, Niger and Togo) and seems to originate from Sudan. Globally we can see that the situation for the serotype O is quite similar to what we observed for the serotype A with only few observed transmissions between the Eastern and western sides of Africa, with Sudan acting as a link between the two sides of Africa.

Phylogeographic tree reconstruction for the serotype SAT2

Due to low posterior probabilities on the location it is quite difficult to tell where the serotype SAT2 emerged, however we estimated that the serotype emerged around 1583 (1722 – 1440 95% HPD) (see fig.1c).

Five geographically defined main clades, with location posterior probabilities above 45% can be observed. The first clade is exclusively composed of sequences from Botswana, Namibia and Zimbabwe and seems to have emerged in the first half of the 19th century. The second clade is composed of Ethiopian, Kenyan, Ugandan and Tanzanian sequences and seems to have emerged at the transition between the 19th and 20th century. The third clade seems to have emerged at the end of the 18th century and is composed of Zimbabwean and all the

South-African sequences. The fourth clade emerged in the first half of the 19th century and is composed of sequences from Botswana, Namibia and Zambia. The last clade is the most more diverse and emerged over the last century and is composed of Eastern, Western and Northern African sequences (Cameroon, Egypt, Ethiopia, Libya, Nigeria and Sudan). Overall, compared to the other serotypes, the SAT2 serotype seems to have emerged first in Southern Africa before moving to other parts of the continent. More over its circulation seem to be more geographically defined with fewer transitions between the sampled countries and with a separation between Southern the other African countries.

Bayesian stochastic search variable selection analysis

Using a Bayesian stochastic search variable selection (BSSVS) analysis we identified well supported rates of transition between the sampled countries. The support for the rates was quantified under the form of a Bayes factor. Globally, the results for the serotypes A and O look quite similar, with Sudan acting as a link between the Eastern, Northern and Western part of the continent (see fig. 2a and 2b). For both serotypes there is a clear transmission route starting from Ethiopia, passing through Kenya to Tanzania. For the SAT2 serotype the situation is quite different to what we observed for serotypes A and O. For the SAT2 serotype most of the observed transmission routes have low BF values (see fig.2c). However, we observe multiple supported transition rates between Soudan, Libya, Nigeria and Cameroon, Sudan acting as link between the Eastern and Central/Western parts of Africa. Finally, all the observed supported rates of transition joining Southern Africa and the other parts of the continent passed through South Africa.

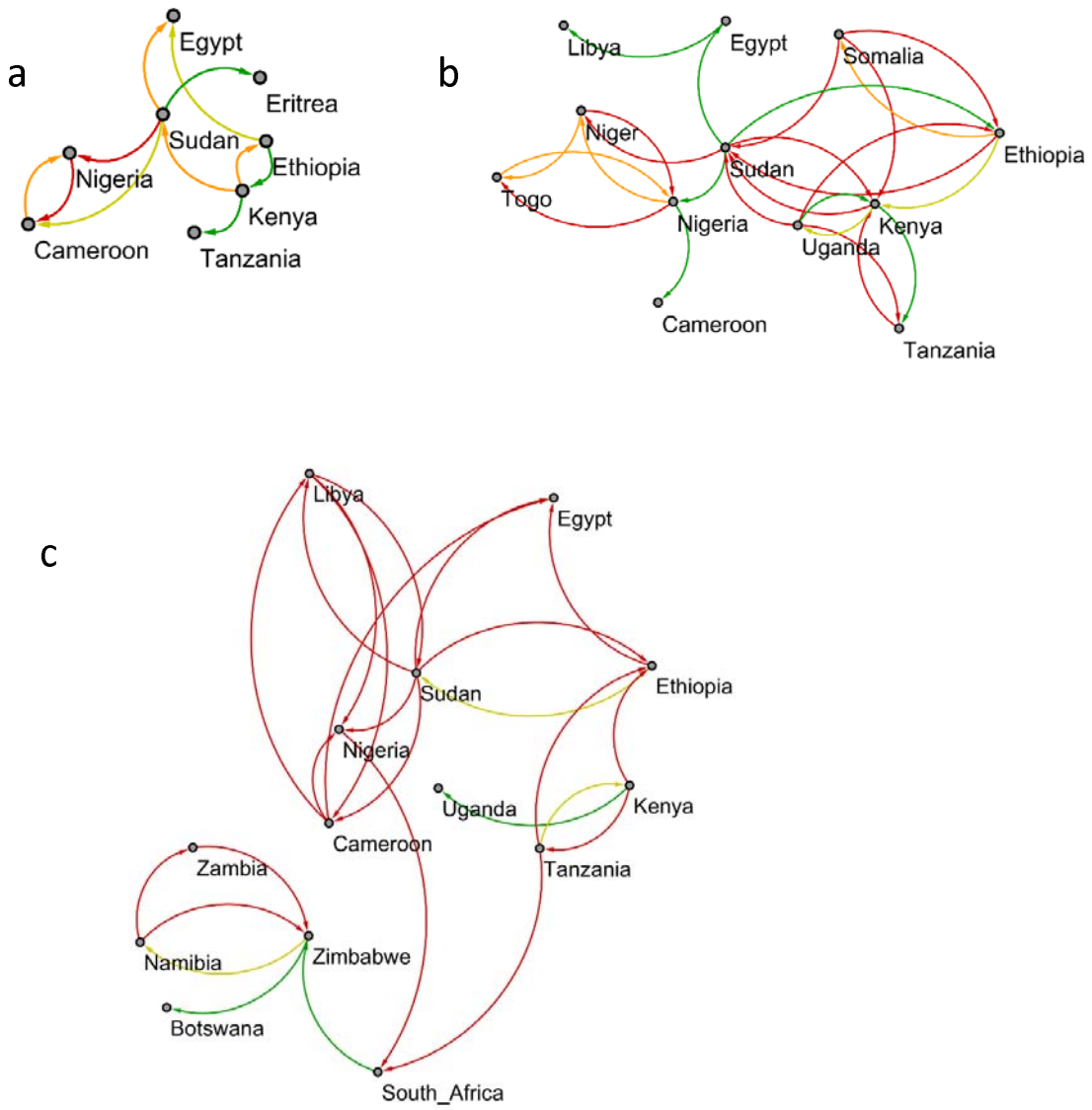

Figure 2 : Outputs of the BSSVS analysis for the three studied FMDV serotypes showing the best supported rates of transition between the sampled countries. The edges colours represent the relative strength by which the rates are supported. a. For FMDV serotype A. b. For FMDV serotype O. c. For FMDV serotype SAT2.

Markov jumps analysis

The BSSVS analysis is complemented by an estimation of the number of transmission events between the different locations using a Markov jump analysis. For both FMDV serotypes A and O we can observe some transmission events starting in Kenya toward other East African countries such as Tanzania, Ethiopia and Uganda. For those two serotypes we also detected jumps from Sudan in the direction of North-Eastern and Western African countries such as Egypt, Eritrea, Cameroon and Nigeria. Therefore, it seems that for those two

serotypes Kenya and Sudan act as sources for the virus, but toward different directions (see supp. table 15 and 16). For the serotype SAT2 the situation is quite different. In this case we observed less transitions events between the different sampled countries (see supp. table 17), most of the observed transitions occurring within Eastern African and Southern Africa with no clear link between them.

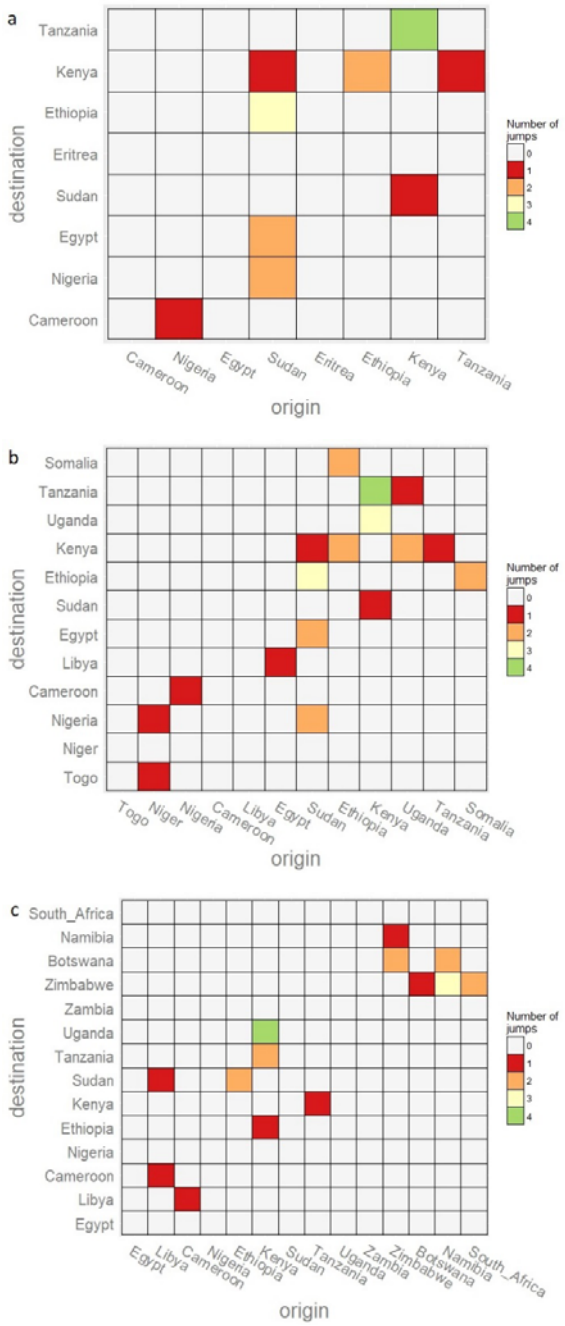

Figure 3: Heatmap showing the number of transitions between the sampled countries for the three studies FMDV serotypes. The graphics are coloured according to the number of observed transitions between countries. a. FMDV serotype A. b. FMDV serotype O. c. FMDV serotype SAT2

Environmental and anthropological effect on the virus circulation

Using a general linear model (GLM) for the discrete approach and the recently developed *SERAPHIM* package [29] for the continuous approach we were able determine the effect that 13 different anthropological and environmental predictors have on the virus diffusion. Therefore, using the output of the previous discrete phylogeographical analysis we isolated a FMDV serotype O monophyletic clade with a time to the most recent common ancestor (TMRCA) below 25 years and a high posterior probability on the location for all its nodes. The selected 46 sequences where originated from Kenya, Tanzania and Uganda.

Table 1 : Environmental predictors for which we assessed the effect on the isolated FMDV serotype O clade 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

Predictors effect in a discrete setting

A generalized linear model (GLM) was used to parametrize the transition rate matrices between the sampled locations as a function of our selected predictors[27]. We observed that the diffusion process was positively affected by the average daily temperature (BF 4), the logarithm of the cattle density (BF 4) and human densities (BF 9) (see Table 2). It was negatively affected by the accessibility (BF 8), the distance between sampled locations (BF 8), the average daily temperature (BF 7) and average amount of precipitation (BF7) per year. To gain a better understanding of the impact of the average temperature and precipitation on the viral circulation we selected different thresholds of precipitation and temperature to parametrize our GLM analysis (see supp. Table 18 and 19). We detected that low precipitation values (< 80 mL/year) were associated with a negative impact on the viral diffusion processes whereas high precipitation was associated with a positive affect on the diffusion process. We also observed that in the case of low temperature (below 22°C) a positive

effect on the circulation could be observed whereas temperatures around 22°C had a negative effect on virus circulation. Temperatures above 24°C seemed again to have a positive impact on the virus spread. It was difficult to distinguish between the effects of accessibility and human density because the two were strongly negatively correlated confounding the analysis (see sup. Table 21.

Predictors effect in a continuous setting

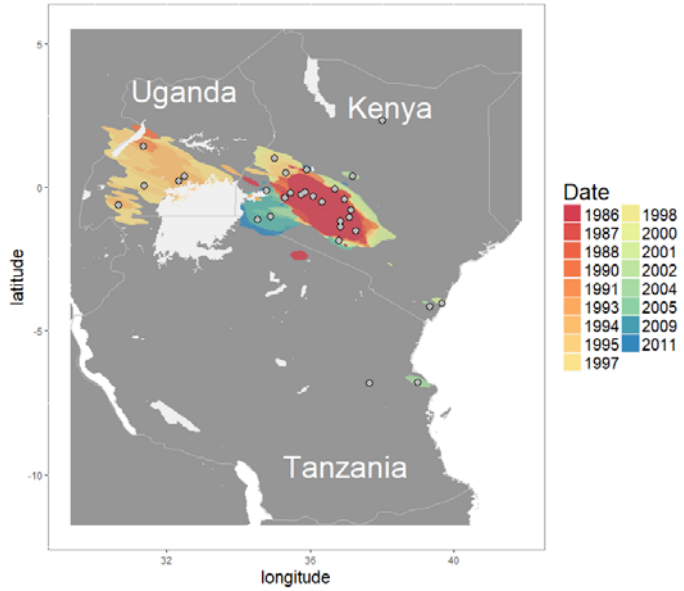

Figure 4: Graphic showing the continuous circulation of the isolated clade of FMDV serotype O, with the sampled locations in grey. The virus movments where reconstructed using a random walk mode with a underlying lognormal distributionl

Using a random walk model, we were able to reconstruct the virus circulation in a continuous setting for the isolated FMDV serotype O. Using the package *SERAPHIM* in R [29] we evaluated the impact of the predictors on the virus circulation and observed a negative influence of the cattle density (BF 3), the presence of cropland (fragmented cropland and pure cropland areas combined) (BF4) and by the presence fragmented cropland (BF6). We were not able to detect a predictor with a positive influence on the diffusion process (see Table 2).

To gain a better understanding of the role of the fragmented crop and cattle density over the spread of the disease we looked at the histogram of frequency for the two predictors at different periods of the epidemic (see supp.fig.5). Overall, we observed an opposite trend for the two predictors in how their density evolve with the course of the infection. For the fragmented crop density, high values of crop densities became more common over the course of the epidemic, the disease moving from areas with low crop densities to areas with high crop densities. For the cattle density the opposite trend was observed with high values of cattle density

more common at the start of the epidemic, the disease starting in an area where cattle densities were high and moving toward areas with lower densities of cattle. To gain a better understanding of the effect that the cattle density had on virus circulation we looked at selected areas above different thresholds of cattle density and used them as inputs in in the *SERAPHIM* package. By doing so we were able to observe that densities of cattle above 125 cattle per square km have the biggest negative impact on the virus circulation (see supp. table 20.

Table 2: Bayes factor values associated with the effect of each predictor on the connectivity between the sampled locations in a discrete or continuous setting. Each predictor raster was use as conductance or resistance to evaluate if the predictor have a positive or negative effect on the viral genetic connectivity.

	Discrete setting		*Continuous setting*	
Predictor	**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)	1	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

DISCUSSION

In this paper we have applied the most recent existing phylogenetic methods on all available African FMDV VP1 sequences for the serotypes A, O and SAT2. Our work has some limitations, especially regarding the limited availability of sequences. Our sampling is obviously unbalanced as it is based on submissions by individual countries or ad hoc research projects and the effect that it has on the results quality is uncertain.

Increasing the number of available FMDV sequences would help to develop models that better represent the circulation of FMDV in Africa and lead to better environmental and anthropological effect estimation.

The estimated substitution rates of $4.67x10^{-3}$, 3.69Ex10-^{3} and $1.1x10^{-3}$ mutations per site per year for the serotypes A, O and SAT2 our results are similar to previous estimates of 4.26×10^{-3}, 3.14×10^{-3} and 1.07×10^{-3} mutations per site per year for the same serotypes[30].

Overall, we observed similar evolutionary patterns for both FMDV serotypes A and O. Both serotypes seem to have emerged in Eastern Africa around 1930. Our results pointed to the possible role of Kenya as a viral source for East African countries and the role of Sudan as a link between East Africa and North-East Africa. The evolution of the SAT2 serotype seems to be quite different. According to our results, this serotype seems to have emerged around 1583. Over the 5 main clades that compose the SAT2 phylogeny only two emerged outside Southern Africa. Those two clades involved Eastern (Kenya, Ethiopia and Tanzania), or Western/Central/Northern African countries (Libya, Nigeria, Cameroon and Egypt) with Sudan being the only country present in both clades. Additionally, these two clades have emerged more recently than those involving Southern African countries (second part of 19th century/early 20th century against late 18th century/early 19th century respectively).

For the FMDV serotypes A and O the observation of well supported rates of viral transmission between Eastern Africa and Western Africa can be explained by the existence of commercial routes between those areas. It is indeed acknowledged that livestock trades play an important role in FMDV circulation in sub-Saharan Africa, through trading routes existing between the Horn of Africa and Eastern Africa, with Sudan acting as key commercial intermediate[9]. Additionally, the existence of a relatively recent common ancestor for the FMDV serotypes A and O is further support for the idea that these serotypes were imported into Africa at the start of the 19th century through livestock trade from Asia and Europe[11].

The SAT2 serotype analysis shows the impact that the African rinderpest epidemics had on FMDV circulation in Africa. Although FMDV was first reported in southern Africa in 1795[31] it likely had coevolved with buffalo over mellenia but the rinderpest epidemic decimated almost all FMDV potential carriers and probably pushed it through a huge bottleneck. It is thought that FMDV re-emerged from wild buffalo population that survived the rinderpest epidemic, before being reported again in 1932 in Southern Africa[11]. Consistent with this hypothesis,

of all the clades present in the reconstructed phylogeny, only those originating from Southern African countries have an TMRCA older than the African rinderpest epidemic. The SAT2 serotype probably spread into non-Southern African countries through infected livestock movements, increasing the virus mobility and explaining the more recent TMRCA and the highest country diversity observed in those clades.

Using both a discrete and continuous framework, we looked at the effect that diverse environmental and anthropological factors had on the diffusion of an isolated FMDV serotype O clade that circulated in Kenya, Uganda and Tanzania. The results of the discrete approaches suggest that the FMDV circulation is facilitated by low average daily temperature (<22℃), high averages precipitations (>80 mL/year) as well as high human and cattle densities. We were also able to observe that the virus circulation was negatively impacted by the accessibility (long travel time needed to join the closest major city) in addition to high daily temperatures and low average precipitations. With lower temperatures and higher humidity values usually associated with a longer virus survivability in the environment[32] our results may suggest a more important role than what was previously believed of the indirect transmission trough viral persistence in the environment for FMV in this region. Additionally, with the viral circulation being positively affected by high cattle and human densities and negatively affected by large accessibility values, it is clear that anthropological activities have an impact on the virus circulation. These observations could be the consequence of infected herds of cattle moving from smaller rural localities toward nearby larger cities with cattle markets[15,33].

Regarding the effect of the different selected predictors on virus circulation in a continuous setting, our results suggest that cattle densities above 125 cattle per km^2 and the presence of cropland (pure cropland or mixed with other types of land) both have a negative impact on virus circulation. Ours results suggest that the virus had difficulties to leave the region located at the root of the tree, where high cattle densities and low crop densities were present and to spread to areas with low cattle densities but high crop densities. However, it is difficult to know exactly whether it is the cattle or crop density that had an impact at the start of the virus propagation due to high correlation of the two variables in the region of origin.

The location uncertainty found at the root of the continuous tree could explain the differences between the discrete and continuous methods in estimating the effect of the cattle density on virus circulation. For our analysis, this uncertainty seems to be translated by the *SEREPHIM* programme as a period where the virus is

almost not moving. This uncertainty seems to drive *SEREPHIM* to the conclusion that the high cattle densities found near the origin of the epidemic are related to this lack of movement and therefore estimate that they have a negative influence on the virus circulation. Although we suspect a link between the cattle density and the location of emergence of the isolated clade, we think that the continuous analysis does not offer the resolution needed to understand that relation. By parameterizing each rate of among-location movement as a function of predictors, the discrete approach seems therefore more appropriate to characterise the environmental and anthropological effect of the virus circulation in this endemic situation.

In conclusion, the reconstituted phylogeographical tree pattern for the FMDV serotypes A and O reflects a situation where the FMDV circulation is mainly driven by commercial exchanges, through pastoral herd movements, and where wildlife seems to have almost no incidence. At the opposite, the observed patterns for SAT2 reflects a situation where cattle movements are responsible for the majority of recent cross border transmission events, while wild buffaloes constitute the original host of the serotype. We observed that indirect transmission through the environment and direct transmission through anthropological activities had a positive impact on the virus circulation in Eastern Africa.

MATERIALS AND METHODS

Data Collection

To obtain a comprehensive genetic dataset we first retrieved all available African FMDV A, FMDV O and FMDV SAT2 genetic sequences in genbank (accessed on the 15/12/2016). From these datasets we selected all VP1 sequences with at least information on the country of sampling and the year of sampling. In total we gather 191 FMDV A, 351 FMDV O and 477 FMDV SAT2 sequences. The sequences were aligned using Multiple Alignment Fast Fourier transformation (MAFFT) [34]. To avoid using sequences with recombination event we used the software RDP4 to remove such sequences [35].

To reduce the effect of potential sampling bias we removed all sequences coming from countries with less than three sequences available and ran a stratified subsampling procedure to allow a maximum of three sequences per country of origin and per month. The spatial coordinates of sampling for each sequence was retrieved using the ggmap package in R and the most precise sampling localisation available for each sequence[36]. The final FMDV A dataset was composed of 107 sequences from eight countries, dates ranging

from 1966 to 2016. The final FMDV O dataset was composed of 192 sequences from 12 countries, dates ranging from 1964 to 2016. The final FMDO dataset was composed of 135 sequences from 15 countries, dates ranging from 1970 to 2015.

Bayesian Evolutionary Inference

Discrete phylogeographical tree inference

Different substitution clock and population evolution models were evaluated by estimating their marginal likelihoods using the Akaike's Information Criterion for MCMC samples (AICM) in Tracer. Ultimately a general-time-reversible (GTR) model with site to site rate variation between two categories was selected as nucleotide substitution model [37] with a Bayesian skygrid population model and a relaxed uncorrelated log-normal molecular clock model were chosen to model the evolution of the FMDA and FMDO serotypes . For the FMD SAT2 serotype a HKY nucleotide substitution model with a constant clock model with a Bayesian skygrid population model were chose to model its evolution[38,39].

Each Markov chain Monte Carlo (MCMC) analysis was performed in BEAST using the BEAGLE library [40]. We first reconstructed the phylogenetic trees for the three studied serotypes. Thereafter to reduce the computation time of the GLM and the spatial diffusion analyses we used a subsets of 1000 trees from the find posterior distribution of tree as an empirical tree distribution. We used TreeAnnotator to summarize maximum clade credibility (MCC) trees and FigTree version 1.4.1 to visualize the annotated trees. [41,42]. The software SPREAD3 and cytoscape were used to identify and visualize the well supported rates of transmission through a Bayes factor test [43].

For the three serotypes we reconstituted the discrete transition events between the different sampled countries through the whole phylogeny using the "mugration model". Therefore, an asymmetric continuous-time Markov chain (CTMC) model with an incorporated Bayesian stochastic search variable selection (BSSVS) was used to determine which set of transition rates sufficiently summarizes the epidemiological connectivity between the countries (Lemey et al. 2009). Moreover, a posterior inference of the complete Markov jump history through the whole genealogy was done, the aim was to quantify state transitions and the time spent in each country by the virus.

Environmental and anthropological effect estimation

Monophyletic lade selection

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

Predictors effect in a discrete setting

A Generalised Linear Model (GLM) extension of the discrete approach was used to test and quantify the positive or negative effect of potential predictors on the viral diffusion process[27]. This model parametrizes the discrete transmission rate between location as a log linear function of the potential predictors. While reconstructing the phylogeographic history the model performs a Bayesian model averaging to determine which combination predictors is the best to explain the spatial diffusion process. The different predictors of FMD diffusion considered in our analyse were: the accessibility to the sampled location, cattle density, crop density, the elevation of the location, the forest density, the human density, the average yearly precipitation, the shrubland area density, the average daily temperature (for the provenance see sup, table 7). For each selected clade different portions of the raster were isolated to only consider the involved countries. To reduce the computational time each raster was aggregated to a resolution of 0.08 by 0.08 corresponding to pixels of approximately 8*8 km.

The circuiscape software was used to estimate values of genetic connectivity between the sampled locations[44]. For each predictor, two values of genetic connectivity were generated, one using the raster as resistance values (impeding the viral circulation) and the other using the raster as conductance value (facilitating the viral circulation). The GLM approach was then performed by parametrizing each of the rates in the CTMC matrix was according to the calculated values of genetic connectivity.

To avoid any auto-correlation between the different predictors the GLM analyse was performed by analysing each predictor with a null predictor corresponding to a ransom raster. The aim was to quantify not only the relative effects that those parameters have on the viral transmission but also their absolute effect.

Predictors effect in a continuous setting

As a complementary approach to the discrete GLM approach, the effect the potential predictors of the diffusion process in continuous space using the SERAPHIM package in R[29]. We first reconstructed the evolution monophyletic clade while inferring its circulation using a random walk model of diffusion. We then used the package SERAPHIM available within R characterise the correlation between the time spend on each phylogeny branch, the covered distance and the environment encountered throughout the whole movement duration[26, 29]. A value of connectivity between the start and end location for each of the phylogeny branch is then assigned according to a "random walk path" model of spatial diffusion according to a raster representing the spatial variation of a potential predictor of viral circulation. The correlation between the time spend on each of the phylogeny branch and the assigned connectivity value is then estimated. The statistical significances of this correlation is then tested using a randomized phylogeny and expressed under the form of a Bayes factor [29].

REFERENCES

1. Alexandersen, S., Zhang, Z., Donaldson, A. I. & Garland, A. J. M. The pathogenesis and diagnosis of foot-and-mouth disease. *J. Comp. Pathol.* **129,** 1–36 (2003).

2. Alexander, D. J. An overview of the epidemiology of avian influenza. *Vaccine* **25,** 5637–5644 (2007).

3. Belsham, G. J. DISTINCTIVE FEATURES OF FOOT-AND-MOUTH DISEASE VIRUS, A MEMBER OF THE PICORNAVIRUS FAMILY; ASPECTS OF VIRUS PROTEIN SYNTHESIS, PROTEIN PROCESSING AND STRUCTURE. *Prog. Biophys. molec. Biol.,* **60,** 241–260 (1993).

4. Fuquay, J. W., Schrijver, R. S. & Vosloo, W. Infectious Diseases: Foot-and-Mouth Disease. *Encycl. Dairy Sci.* 160–167 (2011). doi:http://dx.doi.org/10.1016/B978-0-12-374407-4.00132-1

5. Domingo, E., Baranowski, E., Escarmís, C. & Sobrino, F. Foot-and-mouth disease virus. *Comp. Immunol. Microbiol. Infect. Dis.* **25,** 297–308 (2002).

6. Alexandersen, S. & Mowat, N. Foot-and-mouth disease: host range and pathogenesis. *Curr. Top. Microbiol. Immunol.* **288,** 9–42 (2005).

7. Genevieve V. Weaver, Joseph Domenech, Alex R. Thiermann, and W. B. K. FOOT AND MOUTH DISEASE A LOOK FROM THE WILD SIDE. *J. Wildl. Dis.* **49,** 759–785 (2013).

8. Donaldson, A. I., Alexandersen, S., Sorensen, J. H. & Mikkelsen, T. Relative risks of the uncotrollable (airborne) spread of FMD by different species. *Vet. Rec.* **148,** 602–604 (2001).

9. Di Nardo, a, Knowles, N. J. & Paton, D. J. Combining livestock trade patterns with phylogenetics to help understand the spread of foot and mouth disease in sub-Saharan Africa, the East and Southeast Asia.Middle. *Rev. Sci. Tech.* **30,** 63–85 (2011).

10. Robinson, T. P. *et al. Global livestock production systems*. (2011).

11. Casey, M. B. *et al. Patterns of Foot-and-Mouth Disease Virus Distribution in Africa: The Role of Livestock and Wildlife in Virus Emergence. The Role of Animals in Emerging Viral Diseases* (Elsevier, 2013). doi:10.1016/B978-0-12-405191-1.00002-8

12. Lyons, N. A. *et al.* Impact of foot-and-mouth disease on milk production on a large-scale dairy farm in Kenya. *Prev. Vet. Med.* **120,** 177–186 (2015).

13. Vosloo, W., Bastos, A. D., Sangare, O., Hargreaves, S. K. & Thomson, G. R. Review of the status and control of foot and mouth disease in sub-Saharan Africa. *Rev Sci Tech* **21,** 437–449 (2002).

14. Tekleghiorghis, T., Moormann, R. J. M., Weerdmeester, K. & Dekker, A. Foot-and-mouth Disease Transmission in Africa: Implications for Control, a Review. *Transbound. Emerg. Dis.* **63,** 136–151 (2016).

15. Motta, P. *et al.* Implications of the cattle trade network in Cameroon for regional disease prevention and control. (2017). doi:10.1038/srep43932

16. Vosloo, W., Bastos, A. D. S., Sangare, O., Hargreaves, S. K. & Thomson, G. R. Review of the status and control of foot and mouth disease in sub-Saharan Africa. *Rev. Sci. Tech.* **21,** 437–49 (2002).

17. Bastos, A. D., Boshoff, C. I., Keet, D. F., Bengis, R. G. & Thomson, G. R. Natural transmission of foot-and-mouth disease virus between African buffalo (Syncerus caffer) and impala (Aepyceros melampus) in the Kruger National Park, South Africa. *Epidemiol. Infect.* **124,** 591–8 (2000).

18. Thomson, G. R., Vosloo, W. & Bastos, A. D. S. Foot and mouth disease in wildlife. *Virus Res.* **91,** 145–161 (2003).

19. Ostfeld, R. S., Glass, G. E. & Keesing, F. Spatial epidemiology: An emerging (or re-emerging) discipline. *Trends Ecol. Evol.* **20,** 328–336 (2005).

20. Dion, E., VanSchalkwyk, L. & Lambin, E. F. The landscape epidemiology of foot-and-mouth disease in South Africa: A spatially explicit multi-agent simulation. *Ecol. Modell.* **222,** 2059–2072 (2011).

21. Flood, J. S., Porphyre, T., Tildesley, M. J. & Woolhouse, M. E. The performance of approximations of farm contiguity compared to contiguity defined using detailed geographical information in two sample areas in Scotland: implications for foot-and-mouth disease modelling. *BMC Vet. Res.* **9,** 198 (2013).

22. Kühnert, D., Wu, C.-H. & Drummond, A. J. Phylogenetic and epidemic modeling of rapidly evolving infectious diseases. *Infect. Genet. Evol.* **11,** 1825–1841 (2011).

23. Drummond, A., Pybus, O. G. & Rambaut, A. Inference of Viral Evolutionary Rates from Molecular Sequences. *Adv. Parasitol.* **54,** 331–358 (2003).

24. Pybus, O. G. & Rambaut, A. Evolutionary analysis of the dynamics of viral infectious disease. *Nat. Rev. Genet.* **10,** 540–550 (2009).

25. Lemey, P., Rambaut, A., Drummond, A. J. & Suchard, M. A. Bayesian Phylogeography Finds Its Roots. *PLoS Comput. Biol.* **5,** e1000520 (2009).

26. Lemey, P., Rambaut, A., Welch, J. J. & Suchard, M. A. Phylogeography Takes a Relaxed Random Walk in Continuous Space and Time. *Mol. Biol. Evol.* **27,** 1877–1885 (2010).

27. Lemey, P. *et al.* Unifying Viral Genetics and Human Transportation Data to Predict the Global Transmission Dynamics of Human Influenza H3N2. *PLoS Pathog.* **10,** e1003932 (2014).

28. Dellicour, S., Rose, R. & Pybus, O. G. Explaining the geographic spread of emerging viruses: a new framework for comparing viral genetic information and environmental landscape data. *BMC Bioinformatics* **in press,** 1–12 (2016).

29. Dellicour, S., Rose, R., Faria, N. R., Lemey, P. & Pybus, O. G. SERAPHIM: studying environmental rasters and phylogenetically-informed movements. *Bioinformatics* btw384 (2016). doi:10.1093/BIOINFORMATICS/BTW384

30. Tully, D. C. & Fares, M. A. The tale of a modern animal plague: Tracing the evolutionary history and determining the time-scale for foot and mouth disease virus. *Virology* **382,** 250–256 (2008).

31. Knowles, N. J. FMD history. (1990). Available at: http://www.picornaviridae.com/aphthovirus/fmdv/fmd_history.htm. (Accessed: 25th May 2018)

32. Bartley, L. M., Donnelly, C. A. & Anderson, R. M. Review of foot-and-mouth disease virus survival in animal excretions and on fomites. *Vet. Rec.* **151,** 667–9 (2002).

33. Robinson, S. E. & Christley, R. M. Exploring the role of auction markets in cattle movements within Great Britain. *Prev. Vet. Med.* **81,** 21–37 (2007).

34. Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* (2017). doi:10.1093/bib/bbx108

35. Martin, D. P., Murrell, B., Golden, M., Khoosal, A. & Muhire, B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* **1,** vev003 (2015).

36. Kahle, D. & Wickham, H. ggmap: Spatial Visualization with ggplot2.

37. Tavare, S. Some probabilistic and statistical problems in the analysis of dna sequences. *Lect. Math. Life Sci. (American Math. Soc.* **17,** 57–86 (1986).

38. Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. Relaxed Phylogenetics and Dating with Confidence. *PLoS Biol.* **4,** e88 (2006).

39. Gill, M. S. *et al.* Improving bayesian population dynamics inference: A coalescent-based model for multiple loci. *Mol. Biol. Evol.*

30, 713–724 (2013).

40. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7,** 214 (2007).

41. Rambaut, A. FigTree. (2009). Available at: http://tree.bio.ed.ac.uk/software/figtree/. (Accessed: 25th April 2018)

42. Bouckaert, R. *et al.* BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Comput. Biol.* **10,** e1003537 (2014).

43. Bielejec, F. *et al.* SpreaD3: Interactive Visualization of Spatiotemporal History and Trait Evolutionary Processes. *Mol. Biol. Evol.* **33,** 2167–2169 (2016).

44. McRae, B. H., Dickson, B. G., Keitt, T. H. & Shah, V. B. Using circuit theory to model connectivity in ecology, evolution, and conservation. *Ecology* **89,** 2712–2724 (2008).