Geographic potential of disease caused by Ebola and Marburg viruses in Africa

A. Townsend Peterson\textsuperscript{a,}\textsuperscript{*}, Abdallah M. Samy\textsuperscript{a,b,}\textsuperscript{*}

\textsuperscript{a} Biodiversity Institute, The University of Kansas, Lawrence, KS, 66045, USA
\textsuperscript{b} Faculty of Science, Ain Shams University, Abbassia, Cairo, 11566, Egypt

\textbf{A R T I C L E   I N F O}

Article history:
Received 12 February 2016
Received in revised form 4 June 2016
Accepted 10 June 2016
Available online 14 June 2016

Keywords:
Zaire ebolavirus
Taï Forest ebolavirus
Sudan ebolavirus
Marburg
Africa
Transmission
Ecological niche

\textbf{A B S T R A C T}

Filoviruses represent a significant public health threat worldwide. West Africa recently experienced the largest-scale and most complex filovirus outbreak yet known, which underlines the need for a predictive understanding of the geographic distribution and potential for transmission to humans of these viruses. Here, we used ecological niche modeling techniques to understand the relationship between known filovirus occurrences and environmental characteristics. Our study derived a picture of the potential transmission geography of Ebola virus species and Marburg, paired with views of the spatial uncertainty associated with model-to-model variation in our predictions. We found that filovirus species have diverged ecologically, but only three species are sufficiently well known that models could be developed with significant predictive power. We quantified uncertainty in predictions, assessed potential for outbreaks outside of known transmission areas, and highlighted the Ethiopian Highlands and scattered areas across East Africa as additional potentially unrecognized transmission areas.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Since 1976, scattered human cases and outbreaks have been documented of hemorrhagic fever caused by viruses of the family Filoviridae, caused by the five known Ebola species and the related Marburg viruses (Groseth et al., 2007), which are non-segmented, negative-stranded RNA viruses. These viruses are believed to be hosted in the long term by fruit bats (family Pteropodidae), although full clarity in this issue is largely lacking: that is, solid causal evidence is accumulating regarding the bat Rousettus aegyptiacus as a reservoir for Marburg virus (Amman et al., 2012; Amman et al., 2015; Towner et al., 2009), and a tie of infections in humans to exposure to mines and caves is clear (Peterson et al., 2006). However, evidence regarding the reservoir of Ebola virus is less clear (Pigott et al., 2015). Although compelling temporal and anecdotal links have been pointed out (Leroy et al., 2009), serological evidence has painted a more complex picture, with detections of viruses in multiple bat species, and in regions and under ecological conditions where particular viruses have never been documented (Hayman et al., 2012; Pourrut et al., 2009). Most recently, Ebola surprised the world community with an emergence in Guinea—quite apart from the magnitude of the outbreak, Zaire ebolavirus was unknown in West Africa, as the virus that would have been expected in Guinea was Taï Forest ebolavirus (Bausch and Schwarz, 2014).

The West African Zaire ebolavirus outbreak underlines the need for a predictive understanding of the geographic distribution of these viruses and their potential for transmission to humans across Africa. Although studies on filoviruses almost invariably include a table (e.g., Chippaux, 2014) and/or map (e.g., Polonsky et al., 2014) of known outbreaks, only four studies have gone beyond the occurrences to interpolate or estimate a full potential distribution: early analyses that explored the basic idea (Peterson et al., 2004, 2006) and a recent pair of assessments that took advantage of an additional decade of accumulation of occurrence data (Pigott et al., 2014, 2015). The older analyses are rather dated, in terms of both the occurrence information and the quality of the environmental data and tools for analysis; the new studies, on the other hand, have a number of shortcomings in methodology, which are treated in detail in the Discussion.

This contribution aims to present a more comprehensive view of the geographic potential of Ebola and Marburg viruses known to infect humans, to offer an up-to-date and rigorous view of where these viruses may be found. Specifically, we (1) tested for niche divergence between Ebola species, and consequently treated Ebola species separately in modeling efforts; (2) we developed analyses...
for Marburg virus as well as Ebola; (3) we assessed and addressed pseudoreplication and its effects on model predictions; and (4) we addressed uncertainty in our model outputs. In our model calibration efforts, we took care to control for accessibility of areas, but we also assessed the possibility of long-distance jumps—effectively ‘surprises’ akin to Zaire ebolavirus appearing in West Africa in 2014 (Bausch and Schwarz, 2014).

2. Materials and methods

This paper presents a variety of niche model-based analyses of filovirus (i.e., Ebola and Marburg) potential distributions across Africa. We excluded two filovirus taxa from consideration for lack of sufficient occurrence information (Cuevavirus is known from one site in Spain only) or any occurrence information whatsoever (transmission of Reston ebolavirus from its natural reservoir is not known from any definable location, such that no geographic information is available to us). We are fully cognizant that occurrence information is likely not sufficient to characterize potential distributions of two additional species (Bundibugyo ebolavirus and Taï Forest ebolavirus), but develop preliminary models and associated uncertainty estimates specifically to quantify and demonstrate this insufficiency.

We focused on developing predictions of potential filovirus distributions, but we also carefully avoided overinterpreting our results. That is, we tested niche differentiation before we aggregated multiple species (only Zaire ebolavirus and Sudan ebolavirus were amenable to testing). Analyzing species separately, we were able to develop analyses for three species (Zaire ebolavirus, Sudan ebolavirus, Marburg), but not for the less-well-known Taï Forest ebolavirus and Bundibugyo ebolavirus. We have incorporated several methodological innovations and precautions that are recognized in the biodiversity realm (e.g., model calibration within an accessible area; Peterson et al., 2011), but that have not been adopted widely enough in work to date in the public health field (Peterson, 2014b).

2.1. Input data

We accumulated occurrence data for this study from diverse sources, including our earlier compilation (Peterson et al., 2004; Peterson et al., 2006) and numerous subsequent compilations (Bausch and Schwarz, 2014; Chungula et al., 2014; Chippaux, 2014; Leroy et al., 2009; Mylne et al., 2014; Pourrut et al., 2005; Roddy, 2014); we were unable to use the data set published in association with the recent mapping exercise (Mylne et al., 2014; Pigott et al., 2013) because only the human occurrence data were published for Ebola: animal detection data used in analyses have not, to our knowledge, been made available to the broader community (we did not request them of the authors of those studies, however).

As an independent source, however, we reviewed all posts in the ProMed archives (http://www.promedmail.org/, queries executed 1 November 2014) that included reference to “Ebola,” “Marburg,” or “filovirus;” we included only occurrences cited as “confirmed,” and always verified records against independent information sources. We omitted occurrences detected serologically in bats, in view of the rather odd patterns that such detections have shown—witness, e.g., the detection of Marburg in the Congo Basin (Towner et al., 2007), an entire biome where it has never been found to occur otherwise; evaluation of effects of including bat detection data in niche models was complicated by the fact that the bat detection data have not been made available to the broader community in anything other than summary form (Mylne et al., 2014; Pigott et al., 2014), but we would not incorporate them into analyses until considerably more clarity was available regarding their meaning.

An ideal spatial/environmental analysis of this sort would either incorporate or eliminate effects of spatial autocorrelation of sampling in occurrence data for such models (de Oliveira et al., 2014; Veloz, 2009). For two reasons, however, we did not take such steps. First, the usual concern is that the spatial autocorrelation is in sampling, rather than in the occurrence of the phenomenon itself; if independent occurrences of the phenomenon in question (e.g., filovirus transmission from reservoir to humans) are spatially autocorrelated by their nature (as we suspect is the case with filovirus occurrences, wherein all occurrences of transmission, at least to humans, should be detected and reported), the effects are perhaps of less concern. Second, and more practically, sample sizes even for the three relatively well-characterized filovirus species are nonetheless still small, such that any reduction to remove autocorrelation would be crippling.

For each occurrence in the final list, we used Internet-based electronic gazetteers to add geographic coordinates to the data record (http://www.fallingrain.com/). We assigned a rough estimate of the uncertainty associated with each data record (entering a cave was assigned 200 m uncertainty; a person in a village was assigned 5–10 km; a general description of a region over which a human was infected was assigned 50–170 km, depending on the density of other major landmarks or political regions; Fig. 1). This data set is deposited and fully available in an institutional online repository (http://hdl.handle.net/1808/21024).

A further key element in developing ecological niche models are hypotheses of areas (M) that have been accessible to the species over relevant time periods (Barve et al., 2011); these areas are, in effect, hypotheses of relevant areas that exclude areas never sampled by the species in question. In light of the minimal information available about the natural history and biogeography of filoviruses, we created rather general M hypotheses as $7^\circ$ (~770 km) buffers around known occurrences of each species, but then modified those areas in light of known biogeographic barriers (e.g., lowlands around and northwest of Lake Turkana in northern Kenya separating humid areas of the Ethiopian Highlands from areas to the south and west; see Fig. 1).

We were concerned about the broad spatial autocorrelation patterns in climate data, which lead to rather general models that do not hold much spatial detail, as was the case in our earlier filovirus models (Peterson et al., 2004). At the same time, we resisted the temptation to use the more-detailed data now available from sensors aboard the MODIS satellite: because known filovirus outbreaks date back as far as 1976, using newer imagery would likely lead to inappropriate environmental signals, in view of the large-scale land use changes that have affected Africa.

As a consequence, we used the older Normalized Difference Vegetation Index (NDVI) data from the AVHRR satellite (James and Kalluri, 1994) as an environmental data set in this study (1 km spatial resolution). We used 12 monthly composite NDVI data layers (downloaded, with atmospheric corrections already completed, from UMD, 2001) corresponding to February 1995–January 1996 to capture aspects of land cover and seasonality. These data layers correspond approximately to the midpoint of the time span of filovirus occurrence data used on model development—although this ‘time-averaged’ approach to temporal consistency between occurrence data and environmental data is not satisfactory (e.g., land-use change before or after the date of the environmental data is not taken into account in analyses), a robust ‘time-specific’ methodology is not as-yet available in ecological niche modeling (Peterson et al., 2005). We inspected each layer for artifacts of processing, and concatenated them into ‘raster stacks’ for analysis in QGIS (version 2.4); we then used principal components analysis (PCA) to (1) reduce numbers of dimensions, and (2) create orthogonal (i.e., uncorrelated), highly informative dimensions for analysis. PCA has the added advantage of creating individual components
that respond (at least roughly) to major axes of interest, such as latitude, seasonality, etc.; we believe that the high information content of principal component axes, combined with their orthogonality, more than offsets the loss of interpretability associated with composite axes (Peterson, 2014b). For the initial suite of ecological niche models, we fit PCAs within M hypotheses for each species to be analyzed, as M is the only region of importance to the actual distribution of a species; however, for the final analysis assessing continentwide patterns of suitability and the potential for unexpected emergence events of filovirus species (see below), we used a single PCA over all of Africa, and used the M areas to clip out data sets for analysis so that the environmental data would be on the same axes between the restricted and continentwide datasets.

2.2. Model calibration

We initially evaluated a series of algorithms for use in this study, particularly those facilitated by the openModeller platform (de Souza Muñoz et al., 2011), but we settled on Maxent, version 3.3.3k (Phillips et al., 2006), in light of the convenient bootstrapping features, which allowed us to build in dimensions of uncertainty in our models. We used the jackknife option to identify variables not contributing importantly to model robustness (i.e., extrinsic gain). Once a reduced set of variables was identified, we conducted 10 replicate analyses for each species based on a 50% bootstrap of available occurrence data; we used the median logistic output of replicate analyses as an appropriate summary; although this metric has been interpreted in the form of a probability of occurrence (Phillips et al., 2006), a more conservative interpretation is as a metric of suitability only (Peterson, 2014b), which we follow here. Although we much appreciate the importance of thresholding ecological niche models to avoid overfitting and overinterpretation of model outputs (Peterson et al., 2007), in light of the small sample sizes available to us for all of the species of interest, we generally retained continuous model outputs for interpretation and visualization of suitability patterns.

2.3. Assess niche similarity between species

We tested for departure from niche similarity using the background similarity test developed by Warren et al. (2008), which assesses the similarity or difference of two niche estimates in relation to similarity among large numbers of null replicate models. In effect, the test assesses whether two niche estimates are more different than one would expect given the background similarity of the two accessible regions. For this test, we focused on the
Fig. 2. Effects of pseudoreplication on model outputs for Zaire ebolavirus. The middle map shows the difference between the logistic modeled suitability scores from models based on pseudoreplicated data mimicking recent modeling efforts (Pigott et al., 2014); the upper and lower panels show pixels (in black) in the western and eastern portions of this species’ known distribution in which the pseudoreplicated model presented values out of the range observed in 25 replicate analyses of non-pseudoreplicated data based on random points cast across the uncertainty area for each point.
two Ebola species for which reasonable numbers of occurrences were available (Zaire ebolavirus, with N = 14, and Sudan ebolavirus, with N = 6). We did not develop such tests for Tai Forest ebolavirus and Bundibugyo ebolavirus, or for the two likely-distinct species of Marburg (Peterson and Holder, 2012), which would have much-reduced statistical power owing to low sample sizes.

Random background points were generated from the accessible area for each species, in numbers equal to numbers of real occurrence data available for each species, with 100 replicate samples, and with the least training presence thresholding option specified. These points were used to generate niche models in Maxent, and background models were compared with the real-species model in each replicate; two such comparisons were developed, testing each of the species against the background of the other. We calculated Hellinger’s I indices in R (R Development Core Team, 2013) for all replicates. We used the 5th percentile of each distribution of background similarity values as a critical value in rejecting the null hypothesis of similarity between the two species.

2.4. Pseudoreplication

In the recently published analyses of Ebola potential distributions (Pigott et al., 2014), the authors chose to interpret the many cases of Ebola detected in great apes in Gabon and Republic of the Congo (Sleeman, 2004) as independent transmission events. To assess the effects of this assumption, we compared Table 2 of Pigott et al. (2014) to the occurrence data matrix cited above; we identified single occurrences in the latter that corresponded to multiple occurrences in the former—indeed, among other instances of pseudoreplication, in one case, 21 occurrences in Table 2 of Pigott et al. (2014) related to a single occurrence in our dataset.

Hence, we developed models for Zaire ebolavirus with the single representative (no pseudoreplication) and with multiple representatives (pseudoreplication). To mimic the multiple sites cited for these pseudoreplicates in Table 2 of Pigott et al. (2014), instead of using a single coordinate pair, we used random points cast within the uncertainty radius, which covered a particularly broad area. We then compared the resulting two maps, detecting areas over- and under-emphasized by models including pseudoreplication. We identified areas for which the pseudoreplicated model presented pixel values outside the range of values among 25 replicate models developed without pseudoreplication (see next paragraph); these sites had pseudoreplication effects on model outputs that were particularly pronounced.

2.5. Assess uncertainty

In the case of viral diseases as rare as Ebola and Marburg, it is not sufficient simply to estimate suitability; rather, a crucial element is assessment of uncertainty in model predictions, to the extent possible. Hence, to incorporate uncertainty deriving from positional considerations into our models, we created 25 sets of random points cast one per outbreak within the uncertainty radius of each outbreak (see Fig. 1), following our previous examples (Nakazawa et al., 2010; Peterson et al., 2006). We then developed 25 replicate analyses following protocols listed above, each consisting of 10 bootstrap subsampling analyses. From the 25 medians of the bootstrap replicates, we obtained the maximum and minimum values, and calculated range = maximum – minimum as an index by which to characterize uncertainty. Because the range of values generated in this index will be highly dependent on sample sizes, we use this index principally for comparisons among areas within a particular species, and not for comparisons among species.

2.6. Assess possibility of “surprises”

The known history of Ebola virus transmission events has already seen three events in which the viruses appeared unexpectedly: Zaire ebolavirus and Sudan ebolavirus appearing simultaneously in 1976, Reston ebolavirus apparently coming from the Philippines, and Zaire ebolavirus appearing in Guinea in 2014. In each case, a virus appeared at a place where it was not expected. As a consequence, an important step is to assess the possibility of further such distributional novelties.

We maintained our focus on model calibration within the accessible area M, but then transferred model predictions across all of Africa, which is the appropriate means by which to extend niche models across areas of potential (but not realized) distribution outside of our current concept of the accessible area of each species (Owens et al., 2013). That is, for this analysis, we conducted a principal components analysis across the entire continent, but extracted the portion of Africa within each species’ individual M area for model calibration. Although these principal components were not as specific, the procedure has the advantage of putting the environmental data for within M and across Africa on the same scale. Models were calibrated within M, and then transferred to all of Africa. We assessed the degree to which model transfers would be extrapolative using the MOP metric (Owens et al., 2013).

3. Results

We first assessed ecological niche similarity between Zaire ebolavirus and Sudan ebolavirus. The two null (background similarity) distributions (i.e., Zaire ebolavirus compared to background models for Sudan ebolavirus, and vice versa) had ranges of similarity values of 0.76–0.92 and 0.84–0.98, whereas the observed value was 0.48. As a consequence, we rejected the null hypothesis of niche similarity, in favor of an alternative hypothesis of niche difference. An important implication of this result is that combining two Ebola species with differentiated niches in analyses is not justified; analyzing them together would result in overestimating the breadth of ecological niches.

Assessing effects of pseudoreplication of occurrence data on model outputs, multiple “copies” of occurrence data influenced modeled distributional areas rather dramatically. Specifically, pseudoreplication caused underestimation of suitability of areas in southern Gabon, southern Republic of the Congo, and southeastern Democratic Republic of the Congo (Fig. 2), and overestimation of suitability in parts of the northern Congo Basin and southern Cameroon. These differences caused by pseudoreplicating occurrence data were ‘out of range’ with regard to our models incorporating positional uncertainty, and thereby indicate substantial effects of pseudoreplication, in terms of underestimating suitability, particularly in the southern and southeastern sectors of the Congo Basin.

Modelled suitability and associated uncertainty for each of the 5 African filovirus species are summarized in Figs. 3 and 4. Specifically, Sudan ebolavirus was reconstructed as having a potential distribution focused in northwestern Uganda, northeastern Democratic Republic of the Congo, and southern South Sudan. These predictions were particularly unstable, however, in the Democratic Republic of the Congo; in contrast, areas in Uganda were stably identified as presenting high suitability. Zaire ebolavirus was modeled as having potential distributional areas across the known range of the species, albeit extending more broadly in Cameroon, Ivory Coast, and other countries; predictions in the lower (western) portions of the Congo Basin, such as in Gabon, Equatorial Guinea, and southern Cameroon, showed relatively high uncertainty.
Fig. 3. Modeled suitability for three relatively well-known species of filoviruses: Sudan ebolavirus, Zaire ebolavirus, and Marburg. Left-hand column maps show suitability values, with darker orange indicating more suitable conditions; right-hand column maps show uncertainty in terms of range of median values when models were based on different random representatives of occurrences within specific uncertainty radii (blue = low uncertainty, red = high uncertainty). Location of the maps is indicated via a blue rectangle in an index map for each species. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
Modeled potential distributional areas for Marburg were rather more scattered and diffuse across the accessible area for this species. Our models had some level of difficulty in distinguishing between the more seasonal and scrubby habitats of East Africa and the less seasonal habitats of the Congo Basin; however, uncertainty was focused in those latter regions of the Congo Basin, such that the strength of the prediction of suitability in those regions was unclear. Modeled potential distributional areas for *Taï Forest ebolavirus* and *Bundibugyo ebolavirus* were both diffuse across their respective accessible areas, and all modeled suitable areas were also highly uncertain, such that the distributions of these two species remain poorly characterized. Relationships between uncertainty and suitability were curvilinear and diverse in form in different species (Fig. 5), but clearly reflected effects of low sample sizes of known occurrences of some species (see, in particular, *Taï Forest ebolavirus* in Fig. 5).

Assessing the possibility of unexpected distributional potential as regards Ebola and Marburg virus distributions, some areas emerge as potential distributional areas not within the known distributions of the three species for which sample sizes were sufficient to permit model development (Fig. 6). Specifically, for *Zaire ebolavirus*, two areas could be discerned: southern Ethiopia and the Gulf of Guinea rim of West Africa. For *Sudan ebolavirus* and Marburg, areas in southwestern Ethiopia and East Africa more generally were notable in terms of distributional potential outside of known areas.

4. Discussion

This study updates the current understanding of filovirus geography in Africa. In the 10 years since our original analysis, the number of known outbreaks has approximately doubled, but most have occurred under quite predictable circumstances; however, with augmented input data, as well as improved analytical frameworks, new maps have been developed that should both anticipate future outbreak events and highlight areas of uncertain predictions. Our careful consideration of uncertainty and how its effects percolate through the analysis process, however, indicated that reasonably robust models could be developed only for *Zaire ebolavirus*, *Sudan ebolavirus*, and Marburg; for *Taï Forest ebolavirus* and *Bundibugyo ebolavirus*, occurrence data are simply insufficient to permit rigorous model development at this point.
For Reston ebolavirus, the virus is known only from non-human pri-
mate infections deriving from captive primate colonies, for which no geographic location of transmission from natural reservoir is known; as such, the geographic potential of this species remains entirely obscure.

The four previous analyses of filovirus potential distributional areas (Peterson et al., 2004, 2006; Pigott et al., 2014, 2015) were all lacking in terms of current needs for accurate mapping. The older pair of analyses (Peterson et al., 2004, 2006), developed by Peterson, were based on input occurrence data that included only about half of the records now available, such that those analyses are now quite dated. They were also based on rather dated analytical workflows that did not take advantage of more recent advances in method-
ology (e.g., Barve et al., 2011). Use of GARP (Stockwell and Peters, 1999) in the older studies was an optimal choice at the time, but we were not satisfied by clustered patterns of omission error in initial, experimental models developed in GARP for the present study.

4.1 Filovirus mapping efforts

The more recent analyses (Pigott et al., 2014, 2015) were com-
promised by a series of methodological decisions. These studies opted for use of recent (2000–2012) satellite imagery as environmental data, which sums all land use change since 1976 into a rather dramatic discord between occurrence data and environmental data; our approach minimized this disagreement by choosing
environmental data that coincided with the approximate midpoint of the dates of the occurrence data. As such, yet a further mapping effort taking best advantage of many lessons learned about ecological niche modeling (Peterson, 2014b) was in order.

In these recent analyses (Pigott et al., 2014, 2015), no consideration was paid to accessible areas (i.e., the specific areas that have been available and accessible to species over relevant time periods) in model calibration (Barve et al., 2011). In addition, all Ebola species were lumped together for analysis, notwithstanding possible niche divergence among them (Warren et al., 2008), which was confirmed in our analyses, at least for the two best-known species; this ‘lumping’ of differentiated taxa will generally result in overly-broad estimates of ecological niches. No assessment was made of uncertainty in the predictions that were produced, such that the reader is left only with a prediction, and no idea as to confidence in predictions for any particular species or place.

Finally, and perhaps most importantly, in the occurrence data used to calibrate models in the Ebola-focused paper (Pigott et al., 2014), individual reservoir-to-nonreservoir mammal events were pseudoreplicated: that is, no care was taken to distinguish among multiple versions of the same emergence event in which the virus jumped to nonreservoir mammals, but then may have spread among the nonreservoir animals (gorillas in particular). We suspect that this approach springs from the authors’ focus on the “zoonotic niche” (Pigott et al., 2014, 2015), wherein non-human primates would be considered as zoonotic intermediate hosts, though not necessarily representative of the ecology of the reservoir and transmission from the reservoir to humans. Such errors of lack of care with pseudoreplication and overrepresentation of areas in model calibration have been made in previous analyses (Fichet-Calvet and Rogers, 2009), yet affect the outcomes of niche modeling efforts rather dramatically (Peterson et al., 2014).

4.2. Species accounts

4.2.1 Zaire accounts

This virus is the best-known and best-documented of the filoviruses, with a known range across the humid rainforest belt of Africa. Beyond its known points of occurrence, our models suggested the potential for occurrence farther to the east in the DRC (moderate suitability with moderate uncertainty), and west into Gabon, Equatorial Guinea, and Cameroon (albeit with higher uncertainty). The West African distributional area of this virus did not have a strong signal of high suitability with low uncertainty, which perhaps is a more general quality of this virus (Fig. 5). Beyond its current assumed distributional limits, our model transfer exercise suggested that southern Ethiopia may represent a potential distributional area for this species, although biogeographic barriers isolating Ethiopia from the rest of Africa are significant (Peterson and Martínez-Meyer, 2007).

4.2.2 Sudan accounts

This virus has a restricted distribution, yet its ecological niche signature is the clearest of the viruses analyzed in this study. As can be seen in Fig. 5, the areas predicted as most suitable for this species were identified as such with low uncertainty. That is, areas of northwestern Uganda, particularly between Lake Albert and Lake Victoria, were identified as highly suitable with low uncertainty; northeastern DRC held areas of modeled high suitability, but with higher uncertainty. Model transfer exercises suggested the possibility of suitable areas for this species in Gabon, Republic of the Congo, and in West Africa, although such far-removed areas are of unknown significance.

4.2.3. Marburg

Marburg represents a relatively enigmatic filovirus, in terms of its diversity, ecology, and geography. That is, early Marburg detections were concentrated in East Africa, but one early outbreak was from considerably farther south, in Zimbabwe (Conrad et al., 1978). Peterson et al. (2006) found that the Zimbabwe outbreak matched the niche ‘signature’ of the East African outbreaks, and successfully anticipated the potential for Marburg to emerge elsewhere in southern Africa, including an Angola outbreak (Towner et al., 2006). Our new models identified broad suitable areas across the northeastern DRC, central Uganda, central Kenya, northern Angola, and southwesternmost DRC; many areas of high suitability were identified as such with low uncertainty (Fig. 5), such that these model predictions appear to be useful. In broader projections of the Marburg niche model, broader areas of northeastern DRC, southern Tanzania, and Mozambique were also identified as matching the environmental profile of this virus.

However, Marburg may present more complexity than is currently appreciated: two major lineages exist within the currently-recognized species (Johnson et al., 1996), which occur sympatrically at least in East Africa. Although the two lineages are clearly on separate evolutionary trajectories (Peterson and Holder, 2012), current taxonomic approaches for viruses obfuscate this very-real diversity (Peterson, 2014a). As such, the true diversity of Marburg viruses, and implications for host associations, ecology, and geographic range, must await availability of additional information.

4.2.4 Bundibugyo ebolavirus

This virus was described only recently as a new filovirus (Towner et al., 2008). In spite of minimal occurrence information, model outputs identified areas of the northeastern DRC, southern Uganda, and southwestern Kenya as potentially suitable for this species. However, uncertainty values were high in essentially all of the areas identified as suitable (Fig. 5), such that no clear prediction of suitability was available.

4.2.5 Tai Forest ebolavirus

Tai Forest ebolavirus certainly ranks among the least-known filoviruses, with Bundibugyo ebolavirus (Towner et al., 2008) and a poorly-known filovirus in Europe (Negredo et al., 2011); this knowledge vacuum was reflected in our modeling outputs, as we found no clear signal of suitability, likely a consequence of extremely thin knowledge. In brief, Tai Forest ebolavirus is known best from a single, well-documented case in which a veterinarian performing an autopsy on a dead chimpanzee became infected (Formenty et al., 1999; Le Guenno et al., 1995). Among the many compilations of filovirus occurrences in recent publications (Bausch and Schwarz, 2014; Changula et al., 2014; Chippaux, 2014; Leroy et al., 2009; Mylne et al., 2014; Pourrut et al., 2005; Roddy, 2014), however, only Chippaux (2014) mentions a second detection of what was apparently this species, which was included in our earlier modeling efforts (Peterson et al., 2004). Indeed, Mylne et al. (2014) went so far as to state explicitly that only a single case of this virus was known. The case in question was of a man who crossed from Liberia into Ivory Coast as a refugee, and presented symptoms of viral hemorrhagic fever (W.H.O., 1995a,b); he was diagnosed via serological assays at the Pasteur Institute, and appears to have infected no further individuals.

This 1995 case is of interest for a number of reasons. First of all, it apparently doubles the information available about the spatial distribution of one of the least-known filoviruses, and for that reason alone merits exploration. Second, however, and perhaps more interesting still, is that this 1995 case was apparently diagnosed on serological grounds only (W.H.O., 1995a,b), and apparently has never been sequenced. As no detail is provided regarding the sero-
logical techniques used in the diagnosis, we mention the possibility that this case could thus pertain to either Táu Forest ebolavirus or Zaïre ebolavirus. Perhaps no samples have been retained from this particular case, but it certainly is an intriguing historical point that should not be lost from the view of the field.

4.3 Caveats

The analyses that we have presented herein do come with a number of caveats, of course. The time-averaged nature of our niche modeling analyses represents a rather-general limitation of the correlational niche modeling paradigm as it presently stands—preliminary explorations of the possibility of linking environmental data that are specific to the timing of each individual occurrence datum have been promising (Peterson et al., 2005). However, the methodology has not seen adequate testing as of yet, particularly for species with small sample sizes of occurrence data, such as the species analyzed herein.

The non-analogous environments across some of the broader projections complicate our attempts to assess the potential for surprise outbreaks in the future. That is, our hypotheses of accessible areas were relatively narrow, which affects the generality and reliability of model transfers to other regions (Owens et al., 2013). No solution is available to these complications of extrapolation, so we simply attempt to interpret model transfers under extrapolative conditions with care.

4.4 Conclusions

The geography of the African filoviruses presents several fascinating features on which we comment here. Until recently, the Ebola virus complex appeared to be distributed in a mosaic across Africa, in which no pair of species co-occurred or approached one another; this picture has changed rather drastically in recent years. Particularly intriguing is the concentration of species diversity in East Africa—a small area of southern South Sudan, Uganda, northeastern Democratic Republic of the Congo, and western Kenya, three filovirus species have been documented. Indeed, if recent suggestions about lineage independence and species limits within Marburg (Peterson and Holder, 2012) were heeded, yet a fourth species would be recognized from the region. In contrast, the Congo Basin appears to house but a single species, in spite of many more human and ape infections known, and occurrences of uncertain significance in bats (Towner et al., 2007), particularly in terms of studies based on serological detections.

Although the great bulk of Ebola and Marburg cases that have accumulated in the past decade (i.e., since the initial modeling efforts a decade ago) is centered within the bounds of the then-known and modeled-suitable areas, the 2014 Guinea outbreak reminds scientists that current knowledge is quite incomplete, and that the viruses may emerge in novel areas. Our analysis of the possibility of such events suggests that a clear region of concern could be the highlands of southern Ethiopia (both Ebola and Marburg) and regions of southern Africa such as southern Tanzania, Mozambique, Zambia, and parts of South Africa (Marburg only; Fig. 6). Although appearance of filoviruses in these areas would indeed be surprising, investment of resources in education to avoid medical and public health personnel being caught unawares is probably a good idea.

Acknowledgements

We thank Lindsay Campbell for help with analyses and comments on an early draft of the manuscript, and the Egyptian Fulbright Mission Program (EFMP) for support of AMS during development of this contribution.

References


