Geographic distribution modeling of infectious disease dynamics in space and time

By

Abdallah Mohammed Samy Esmail Gad

Submitted to the graduate degree program in Ecology and Evolutionary Biology and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

_____
Chairperson A. Townsend Peterson

_____
Jorge Soberón

_____
Benjamin A. Sikes

_____
Folashade Benette Agusto

_____
Joane Nagel

Date Defended: May 5, 2016

The Dissertation Committee for Abdallah Mohammed Samy Esmail Gad
certifies that this is the approved version of the following dissertation:

Geographic distribution modeling of infectious disease dynamics in space and time

_____

Chairperson A. Townsend Peterson

Date approved: May 5, 2016

Abstract

The geographic distribution of infectious diseases has received considerable attention after several dramatic emergence events around the world. Here, I took the full advantages of several approaches available in a single toolbox to examine geographic distribution and spread of several neglected and zoonotic diseases across the world. These approaches included geographic information system, remote sensing, ecological niche modeling, and phylogeography of disease outbreaks. The results assessed and evaluated several diseases based on their public health importance, data availability, and geographic dimension. These diseases included major neglected tropical diseases of potential public health worldwide (e.g. mycetoma, and leishmaniasis), zoonosis (e.g. Rift Valley Fever), and livestock diseases (e.g. Bluetongue). In 2013, the World Health Organization (WHO) recognized mycetoma as one of the neglected tropical conditions due to the efforts of the mycetoma consortium. This same consortium formulated knowledge gaps that require further research. One of these gaps was that very few data are available on the epidemiology and transmission cycle of the causative agents. Previous work suggested a soil-borne or *Acacia* thorn-prick-mediated origin of mycetoma infections, but no studies have investigated effects of soil type and *Acacia* geographic distribution on mycetoma case distributions. In chapter 1, the study mapped risk of mycetoma infection across Sudan and South Sudan using ecological niche modeling (ENM). I developed ENMs based on case occurrences, and digital GIS data layers summarizing soil characteristics, land-surface temperature, and greenness indices to provide a rich picture of environmental variation across Sudan and South Sudan. ENMs were calibrated in known endemic districts and transferred countrywide; model results suggested that risk is greatest in an east-west belt across central Sudan. Visualizing ENMs in environmental dimensions, mycetoma occurs under diverse

environmental conditions. The study also compared niches of mycetoma and *Acacia* trees, and could not reject the null hypothesis of niche similarity. This study revealed contributions of different environmental factors to mycetoma infection risk, identified suitable environments and regions for transmission, signaled a potential mycetoma-*Acacia* association, and provided steps towards a robust risk map for the disease.

In chapter 2, I studied another neglected tropical disease in Libya where political instability prevent active surveillance of cutaneous leishmaniasis (CL). CL ranks among the tropical diseases least known and most neglected in Libya. World Health Organization reports recognized associations of *Phlebotomus papatasi*, *Psammomys obesus*, and *Meriones* spp., with transmission of zoonotic cutaneous leishmaniasis (ZCL; caused by *Leishmania major*) across Libya. Here, the study map risk of ZCL infection based on occurrence records of *L. major*, *P. papatasi*, and four potential animal reservoirs (*Meriones libycus*, *Meriones shawi*, *Psammomys obesus*, and *Gerbillus gerbillus*). Ecological niche models identified limited risk areas for ZCL across the northern coast of the country; most species associated with ZCL transmission were confined to this same region, but some had ranges extending to central Libya. All ENM predictions were significant based on partial ROC tests. As a further evaluation of *L. major* ENM predictions, the study compared predictions with 98 additional independent records provided by the Libyan National Centre for Disease Control (NCDC); all of these records fell inside the belt predicted as suitable for ZCL. The study tested ecological niche similarity among vector, parasite, and reservoir species and could not reject any null hypotheses of niche similarity. Finally, I tested among possible combinations of vector and reservoir that could predict all recent human ZCL cases reported by NCDC; only three combinations could anticipate the distribution of human cases across the country.

Further in chapter 3, I developed a comprehensive occurrence data set to map the current distribution, estimate the ecological niche, and explore the future potential distribution of BTV globally using ecological niche modeling and based on diverse future climate scenarios from general circulation models (GCMs) for four representative concentration pathways (RCPs). The broad ecological niche and potential geographic distribution of BTV under present-day conditions reflected the disease's current distribution across the world in tropical, subtropical, and temperate regions. All model predictions were significantly better than random expectations. As a further evaluation of model robustness, I compared our model predictions to 331 independent records from most recent outbreaks from the Food and Agriculture Organization Emergency Prevention System for Transboundary Animal and Plant Pests and Diseases Information System (EMPRES-i); all were successfully anticipated by the BTV model. Finally, I tested ecological niche similarity among possible vectors and BTV, and could not reject hypotheses of niche similarity. Under future-climate conditions, the potential distribution of BTV was predicted to broaden, especially in central Africa, United States, and western Russia.

Finally, in chapter 4, I used phylogenetic analyses to understand the demographic history of RVFV populations, using sequence data from the three minigenomic segments of the virus. I used phylogeographic approaches to infer RVFV historical movement patterns across its geographic range, and to reconstruct transitions among host species. Results revealed broad circulation of the virus in East Africa, with many lineages originating in Kenya. Arrival of RVFV in Madagascar resulted from three major waves of virus introduction: the first from Zimbabwe, and the second and third from Kenya. The two major outbreaks in Egypt since 1977 possibly resulted from a long-distance introduction from Zimbabwe during the 1970s, and a single introduction took RVFV from Kenya to Saudi Arabia. Movement of the virus between

Kenya and Sudan, and CAR and Zimbabwe was in both directions. Viral populations in West Africa appear to have resulted from a single introduction from Central African Republic. Finally, host transition analysis identified both humans and livestock as natural hosts of RVFV. The overall picture of RVFV history is thus one of considerable mobility, and dynamic evolution and biogeography, emphasizing its invasive potential, potentially more broadly than its current distributional limits.

The results raised by all these analyses offered the potential capacity of ecological modeling and phylogeographic approaches to understand the potential distribution and spread of different disease systems and open the possibilities for their applications in understanding disease epidemiology for surveillance and control efforts of several other disease systems emerged recently across the world.

Acknowledgments

I would like to express my deep gratitude and sincere acknowledgment to all of those who helped me throughout my study at the University of Kansas. All thanks go to my advisor Town Peterson for his continuous support. Special thanks extend also to all other members of my PhD committee; Jorge Soberón, Joane Nagel, the late Val H. Smith, Benjamin A. Sikes, and Folashade Agusto for their times and advices throughout the whole process. They were not just advisers to guide all of my entire work through the time of the study, they also stood as family, friends, and colleagues here in Lawrence. Town Peterson spent a lot of his time offering helps and several skills on both of my academic and research trajectories. Town also provided advices to follow in my future career and to collaborate with colleagues from across the world.

With Town Peterson and Jorge Soberón, my committee was very fruitful in discussion, advices, and recommendations. Joane Nagel helped me during all work and provided advices on how to deal with local communities in study areas that I visited. I am still remember her advices before my travel to Egypt, and Sudan. Comments from the late Val H. Smith, Benjamin A. Sikes, and Folashade Agusto further improved the work. All of my committee members were great advisers to this study and provided guidance through the entire work.

I am grateful to my coauthors for their continuous support during the study and for the scientific knowledge I gained by collaborating with them. Specifically, I like to thank Wendy W. J. van de Sande from Erasmus Medical Center, Department of Medical Microbiology and Infectious diseases, Rotterdam, The Netherlands, Matthew D. Hall from the Department of Infectious Disease Epidemiology, Imperial College London, London, Ahmed Hassan Fahal from the Mycetoma Research Centre, University of Khartoum, Khartoum, Sudan, Mohamad A. Kenawy from the Department of Entomology, Faculty of Science, Ain Shams University, Egypt,

Badereddin B. Annajar from the Libyan National Centre for Disease Control (NCDC), Tripoli, Libya, and Samia Boussaa from Laboratory of Ecology and Environment (URAC 32, CNRST; ERACNERS 06), Faculty of Sciences Semlalia, Cadi Ayyad University, Marrakesh, Morocco, and Institut Supérieur des Professions Infirmières et des Techniques de Santé (ISPITS), Ministry of Health, Marrakesh, Morocco.

KU-ENM group were the source of motivation by their fruitful discussion, ideas, and questions. They developed my quantitative skills during my program at KU. It was of great pleasure to work with all of them; Thanks to Cori, Erin, Hannah, Andres, Narayani, Vijay, Chris, Kate, Lindsay, Marianna, Katie, Abdulghafar, Karen, and Ali. Special thanks go to my best friends here; Abdelghafar Alkeish, Ahmed Alenazi, and Katie Allen. They were very kind and wonderful people; I hoped they came early, so, we can continue working together for longer time.

My acknowledgment go also to my best friends who came to Peterson laboratory as visiting researchers, postdocs, and visiting students during the last four years including Qiao Huijie, Guto Hashimoto Mendonça, Vivian Ribeiro, Carlos Yañez-Arenas, Alondra Encarnación-Luévano, and Gengping Zhu.

All members of family are much appreciated for decreasing the burden of travel outside the country and for their continuous support. Special thanks to my mum, dad, brother, sisters, and to my wife for their support and patience during the study. They inspired me with the best beautiful family. Thanks also go for Rosi Peterson and Town Peterson who became members of my family since my arrival to Lawrence and for the rest of my life. They were better than a family and I never feel that I am away from my country when I look for the face of this nice couple. Thanks a lot for both Town and Rosi for everything they did to me in this new world.

Thanks for support from the Egyptian Fulbright Mission Program, American-Mideast Educational and Training Services (Amideast), WHO-TDR small grant, American Society of Tropical Medicine and Hygiene Young Investigator Award, PLOS Global Participation Initiative (PLOS GPI) Fund, University of Kansas Center for Research (KUCR) Travel Fund, BI's Panorama program, Department of Ecology and Evolutionary Biology, BI's Charles Fund, Leona Galutia Burt Memorial Fund, Sudanese American Medical Association (SAMA) grant, Ain Shams University, and KU open Access Fund. All of these organizations contributed funding through all stages of study, presentation, and publications.

Dedication

I dedicate my dissertation work to my family and many friends. A special feeling of gratitude to my loving parents, sisters, brother, and my wife.

I also dedicate this dissertation to my academic advisers, and friends who have supported me throughout the whole process since my arrival to USA. I will always appreciate all they have done, especially A. Townsend Peterson for helping me to develop my skills for the many hours of manuscript reading and advices to improve my work.

I dedicate this work and give special thanks to my wonderful mum Rosi Peterson, and my wonderful colleagues Soliman Alghafri, and Abdelghafar Alkeish for being there for me throughout the entire doctoral program.

Table of Contents

List of Figures

Chapter 1: Mapping the potential risk of mycetoma infection in Sudan and South Sudan using ecological niche modeling

Chapter 3: Climate change influences on the global potential distribution of bluetongue virus

Chapter 4: Phylogeography and host transitions of Rift Valley Fever virus in Africa and the Arabian Peninsula

List of Tables

List of supporting Information

Chapter 3: Climate change influences on the global potential distribution of bluetongue virus

Chapter 4: Phylogeography and host transitions of Rift Valley Fever virus in Africa and the Arabian Peninsula

Introduction

Infectious diseases are one of the leading causes of death globally, following cardiovascular disease [1]; however, the mechanisms underlying their emergence are still not fully understood [2]. Recently, several diseases have emerged with no details known for their ecology, and even their primary reservoirs are still debated 5 years post-emergence (e.g. coronavirus; [3]). Regardless, the transmission dynamics of pathogens, host and vector distributions are highly sensitive to environmental factors and particularly to dynamics of climate and land use (see, e.g. [4]; [5]), and their emergence can be anticipated in light of certain environmental trends (e.g. land use changes, global climate change; [6]. On the other hand, pathogenic and parasitic organisms become a common and integral part of ecosystems, and also influence the abundance of wild host populations, and thus can cause shift in host ranges, diminished ranges or even extinctions of their hosts, and serve as drivers of disease evolution [7].

Studies of disease ecology require interdisciplinary efforts to explain the temporal shifts in the underlying environmental or demographic state. For example, a Rift Valley Fever (RVF) outbreak was first identified near Lake Naivasha in the Rift Valley area of Kenya in 1931 [8] and then in Egypt in 1977 before the major outbreaks of RVF in 1997-1998, and 2000 again in Kenya, Somalia, and Tanzania, and Saudi Arabia [9]. The geographic distribution of RVF was known only in Africa till the first outbreak identified outside Africa in Saudi Arabia [10]. RVF patterns suggest that importation of infected animals to the new outbreak sites raises the most significant risk for future spread of RVF outside Africa. It is therefore reasonable to estimate the possibility that RVF introductions to Americas can arise from the increasing of demographic and economic activities between US and the other countries of RVF epidemics [11]. Two important

risk factors can be tested for a successful introduction of RVF to US: 1) presence versus absence of disease vectors, possibility of introducing infected hosts, or vectors to America, and 2) the environment affecting disease dynamics. Unfortunately, a single or small population of infected vector can secure the transmission cycle if successfully passed to local animal hosts and competent domestic vectors [12].

Globally, several diseases were transmitted via vectors. These diseases caused historical epidemics across the world, for example, RVF, leishmanisis, plague, malaria, and arboviruses [13]. For example, there was a concern for the diverse arboviral zoonoses in which *Aedes albopictus* might participate in its new environments in America [14]. However, these mechanisms can occur and explain the dynamics of disease systems in only one side of their transmission cycle (i.e. vector), but, ecologists and epidemiologists should be interested to investigate all organisms involved in the disease transmission cycle, the objective that we are interested to explore at one of our analyses. On the other hand, the transmission cycle may be interrupted due to the absence of the primary host for either conservation reasons (i.e. range shifts in response to climate, endangered species, went extinct). For example, a study revealed that the acute epidemic of Ebola virus (EBOV) in an extensive and complex ecosystem such as Central African forest may quickly run out of susceptible primates and duikers [15]; [16]). A compelling circumstantial evidence was identified that the massive chimpanzee-population decline attributed to EBOV in the Congo-Gabon region [16]; [17]). However, EBOV may threat the chimpanzee population from the conservation mind but also give signals for the changes in the primary host and raises the concerns from being a spillover to EBOV from epidemiology view. This spillover is also reported for the emergence of a filovirus in a swine as a new mammalian host [18].

Here, I developed some ecological hypotheses for different diseases of interest to the public. These hypotheses considered important approaches to a single toolbox; geographic information system, remote sensing, ecological niche modeling, and phylogeography of outbreaks. Previous phylogeographic approaches explained the historical dispersal patterns of viruses based on model-free heuristic approaches that provide little insight into the temporal setting of the spatial dynamics [19]. Several studies enabled the reconstruction of temporal dispersal patterns in disease dynamics: avian influenza [19], foot and mouth disease [20], and rabies [21]. Here, I took the full advantages of these integrated tools to test several ecological, and phylogeographic questions associated with the disease transmission system. In this study, I assessed and evaluated several disease models based on public health importance, data availability, and geographic dimension.

This dissertation has a series of goals, all centered on the idea of understanding the ecology, geography, and phylogeography of several disease transmission models across the world, with special interest in the African continent. This offer full four chapters that cover different aspects of disease ecology and dynamics using the available toolbox and disease mapping tools.

In chapter 1, I investigated risk factors associated with mycetoma infections in Sudan using ecological niche modeling (ENM), integrating mycetoma case records, *Acacia* records, and geospatial data summarizing soil, land-surface temperature, and greenness. ENMs calibrated in endemic districts were transferred across Sudan, and suggested that greatest risk was in a belt across central Sudan. Mycetoma infections occur under diverse environmental conditions; I found significant niche similarity between *Acacia* and mycetoma. Model predictions were amply

corroborated by a preliminary assessment of a much larger mycetoma case-occurrence data base. Our results revealed contributions of different environmental factors to mycetoma risk, raised hypotheses of a causal mycetoma-*Acacia* association, and provide steps towards a robust predictive risk map for the disease in Sudan.

In chapter 2, I used ecological niche model as a tool for risk-mapping of both ZCL cases and distributions of associated species. This model offer a challenge to ecological niche modeling considering the very complex transmission system of leishmaniasis. Our models were able to anticipate areas of highest risk with statistical significance, lending confidence that they were successful in identifying areas of transmission risk. Zoonotic cutaneous leishmaniasis (ZCL) represents a major public health problem in North Africa where *Leishmania major* is the potential etiological agent associated with all ZCL cases. In many countries across North Africa, *L. major* is transmitted by the sand fly *Phlebotomus papatasi*, with rodents as likely reservoir hosts. In Libya, ZCL cases are underestimated for lack of reporting, insufficient information about the distribution of ZCL, and interactions between local environmental conditions and different disease components. This situation worsened with recent political and socio-economic changes in the country, with expansion and rapid increases in numbers of cases across the country. For management and planning of leishmaniasis control, predicting the potential geographic distribution of risk of infection with the disease is important to guide such programs.

In chapter 3, I developed a comprehensive occurrence data set to map the current distribution, estimate the ecological niche, and explore the future potential distribution of BTV globally using ecological niche modeling and based on diverse future climate scenarios from general circulation models (GCMs) for four representative concentration pathways (RCPs). The

broad ecological niche and potential geographic distribution of BTV under present-day conditions reflected the disease's current distribution across the world in tropical, subtropical, and temperate regions. All model predictions were significantly better than random expectations. As a further evaluation of model robustness, I compared our model predictions to 331 independent records from most recent outbreaks from the Food and Agriculture Organization Emergency Prevention System for Transboundary Animal and Plant Pests and Diseases Information System (EMPRES-i); all were successfully anticipated by the BTV model. Finally, I tested ecological niche similarity among possible vectors and BTV, and could not reject hypotheses of niche similarity. Under future-climate conditions, the potential distribution of BTV was predicted to broaden, especially in central Africa, United States, and western Russia.

In chapter 4, I applied phylogographic approaches to understand the demographic history of RVFV populations, using sequence data from the three minigenomic segments of the virus. These analyses enabled understanding the overall picture of virus mobility and host transition across the virus geographic range.

Ecological and geographic spread of these diseases became one of priorities to understand emergence of disease outbreaks and identify research priorities for surveillance and control in endemic areas across the world. I focused our study on these disease systems in terms of the international demand to understand the ecological and geographic dimensions of their spread. The overall study was translated into actions in several control programs across the world, for example, our mapping efforts to mycetoma are now used by the international mycetoma consortium. I also identified priorities of sampling of CL in East Libya which had never considered as a target for surveillance and control programs across the country. These

analyses revealed also the importance of considering sampling from other organisms in the cycle to better guide the control program across the country. The study also make the global map of bluetongue virus available to identify the areas of risk across the world and inferred possible future shifts for the disease transmission using different emission scenarios available from the most recent future climate data. Finally, the phylogeographic analyses of RVFV assisted to understand the dynamics and possible spread of the virus across Africa and the Arabian Peninsula. It identified also the possible host transitions of RVFV.

References:

1. Fauci AS (2005) Emerging and reemerging infectious diseases: the perpetual challenge. Acad Med 80 (12): 1079-1085.
2. Daszak P, Zambrana-Torrelio C, Bogich TL, Fernandez M, Epstein JH, et al. (2013) Interdisciplinary approaches to understanding disease emergence: the past, present, and future drivers of Nipah virus emergence. Proc Natl Acad Sci USA 110 (1) 1: 3681-3688.
3. de Sousa R, Reusken C, Koopmans M (2014) MERS coronavirus: data gaps for laboratory preparedness. J Clin Virol 59 (1): 4-11.
4. Peterson AT, Martínez-Campos C, Nakazawa Y, Martínez-Meyer E (2005) Time-specific ecological niche modeling predicts spatial dynamics of vector insects and human dengue cases. Trans R Soc Trop Med Hyg 99 (9): 647-655.
5. Samy AM, Campbell LP, Peterson AT (2014) Leishmaniasis transmission: distribution and coarse-resolution ecology of two vectors and two parasites in Egypt. Rev Soc Bras Med Trop 47 (1): 57-62.

6. Peterson AT (2009) Shifting suitability for malaria vectors across Africa with warming climates. BMC Infect Dis 9: 59.
7. Hudson PJ, Rizzoli, A., Grenfell, B. T., Heesterbeek, H., Dobson, A. P. (2002) The Ecology of Wildlife Disease. Oxford, UK, Oxford Press.
8. Daubney R, Hudson JR, Garnham PC (1931) Enzootic hepatitis or Rift Valley Fever. An undescribed virus disease of sheep cattle and man from East Africa. J Pathol Bacteriol 34 (4): 545-579.
9. Nderitu L, Lee JS, Omolo J, Omulo S, O'Guinn ML, et al. (2011) Sequential Rift Valley Fever outbreaks in eastern Africa caused by multiple lineages of the virus. J Infect Dis 203 (5): 655-665.
10. Madani TA, Al-Mazrou YY, Al-Jeffri MH, Mishkhas AA, Al-Rabeah AM, et al. (2003) Rift Valley Fever epidemic in Saudi Arabia: epidemiological, clinical, and laboratory characteristics. Clin Infect Dis 37 (8): 1084-1092.
11. Hartley DM, Rinderknecht JL, Nipp TL, Clarke NP, Snowder GD (2011) Potential effects of Rift Valley Fever in the United States. Emerg Infect Dis 17 (8): e1.
12. Rolin AI, Berrang-Ford L, Kulkarni MA (2013) The risk of Rift Valley Fever virus introduction and establishment in the United States and European Union. Emerg Microbes Infect 2: e81.
13. Lounibos LP (2002) Invasions by insect vectors of human disease. Annu Rev Entomol 47: 233-266.
14. Mitchell CJ (1991) Vector competence of North and South American strains of *Aedes albopictus* for certain arboviruses: a review. J Am Mosq Control Assoc 7: 446-451.

15. Walsh PD, Abernethy KA, Bermejo M, Beyers R, De Wachter P, et al. (2003) Catastrophic ape decline in western equatorial Africa. Nature 422: 611-614.
16. Breed A, Plowright R, Hayman DS, Knobel D, Molenaar F, et al. (2009) Disease Management in Endangered Mammals. Springer Japan. pp. 215-239.
17. Leroy EM, Telfer P, Kumulungui B, Yaba P, Rouquet P, et al. (2004) A serological survey of Ebola virus infection in central African nonhuman primates. J Infect Dis 190 (11): 1895-1899.
18. Marsh GA, Haining J, Robinson R, Foord A, Yamada M, et al. (2011) Ebola Reston virus infection of pigs: clinical significance and transmission potential. J Infect Dis 204 (3): S804-809.
19. Lemey P, Rambaut A, Drummond AJ, Suchard MA (2009) Bayesian Phylogeography Finds Its Roots. PLoS Comput Biol 5: e1000520.
20. Hall MD, Knowles NJ, Wadsworth J, Rambaut A, Woolhouse ME (2013) Reconstructing geographical movements and host species transitions of foot-and-mouth disease virus serotype SAT 2. mBio 4: e00591-00513.
21. Kuzmina NA, Lemey P, Kuzmin IV, Mayes BC, Ellison JA, et al. (2013) The phylogeography and spatiotemporal spread of south-central Skunk rabies virus. PloS one 8: e82348.

Chapter 1: Mapping the potential risk of mycetoma infection in Sudan and South Sudan using ecological niche modeling[1]

[1] Samy AM, van de Sande WW, Fahal AH, Peterson AT (2014). Mapping the potential risk of mycetoma infection in Sudan and South Sudan using ecological niche modeling. PLoS Negl Trop Dis. 8(10):e3250. doi: 10.1371/journal.pntd.0003250

Abstract

In 2013, the World Health Organization (WHO) recognized mycetoma as one of the neglected tropical conditions due to the efforts of the mycetoma consortium. This same consortium formulated knowledge gaps that require further research. One of these gaps was that very few data are available on the epidemiology and transmission cycle of the causative agents. Previous work suggested a soil-borne or *Acacia* thorn-prick-mediated origin of mycetoma infections, but no studies have investigated effects of soil type and *Acacia* geographic distribution on mycetoma case distributions. Here, we map risk of mycetoma infection across Sudan and South Sudan using ecological niche modeling (ENM). For this study, records of mycetoma cases were obtained from the scientific literature and GIDEON; *Acacia* records were obtained from the Global Biodiversity Information Facility. We developed ENMs based on digital GIS data layers summarizing soil characteristics, land-surface temperature, and greenness indices to provide a rich picture of environmental variation across Sudan and South Sudan. ENMs were calibrated in known endemic districts and transferred countrywide; model results suggested that risk is greatest in an east-west belt across central Sudan. Visualizing ENMs in environmental dimensions, mycetoma occurs under diverse environmental conditions. We compared niches of mycetoma and *Acacia* trees, and could not reject the null hypothesis of niche similarity. This study revealed contributions of different environmental factors to mycetoma infection risk, identified suitable environments and regions for transmission, signaled a potential mycetoma-*Acacia* association, and provided steps towards a robust risk map for the disease.

Introduction

Mycetoma is a chronic, devastating, inflammatory disease of the subcutaneous tissues that spread to involve the skin, deep structures and bones, and is characterized by deformity, destruction and disability especially in late stages [1-3]. Etiological agents are identified by culturing their characteristic compact mycelial grains [4,5]. The infection most often affects the lower extremities of individuals living in developing tropical and subtropical countries [6]. Two forms of mycetoma have been identified [3,7]: actinomycetoma caused by a group of filamentous bacteria, and eumycetoma caused by any of 30-50 species of hyaline and pigmented fungi [4,8-11].

The organisms causing mycetoma are geographically distributed worldwide, but are particularly common in tropical and subtropical areas, in the so-called 'mycetoma belt,' which includes Mexico, Venezuela, Mauritania, Senegal, Chad, Ethiopia, Sudan, Somalia, Yemen, and India [11]. The incidence and geographic distribution of mycetoma are underestimated, as the disease is usually painless and slowly progressive, such that it is presented to health centers only in late disease stages by most of patients; it is not a reportable disease [12-14]. Mycetoma is a socioeconomically biased disease, and typically appears in low-income communities with poor hygiene; for example, agricultural laborers and herdsmen appear worst affected [15,16]. Studies revealed that minor traumas can allow pathogens to enter the skin from the soil [7], or through *Acacia* thorns, to the point that *Acacia* thorns have been found embedded in mycetoma lesions during surgery [4,17]. Fungal infections responsible for eumycetoma in Sudan are predominantly caused by *Madurella mycetomatis* [4].

Studies to date suggest a soil-borne or thorn-prick-mediated origin of mycetoma infections [4], having demonstrated *M. mycetomatis* DNA on *Acacia* thorns and in soil samples [4]. Although prevailing thought is that the soil is the ultimate reservoir for mycetoma infections, attempts to culture the fungus from soil samples have failed [4,14]. A more recent study suggested that cattle dung may play a significant role in the ecology of *Madurella*, based on the observation that *M. mycetomatis* is phylogentically closely related to dung-inhabiting fungi [18].

Mycetoma ranks among the most neglected diseases worldwide, to the point that it was omitted even by major neglected tropical disease (NTD) initiatives across the globe [19-21]. Recently, mycetoma was added to the WHO's list of NTD priorities [11]. The known geographic distribution of mycetoma etiological agents shows intriguing variation with respect to environmental factors [22]: they occur in arid areas with a short rainy season, and extreme conditions have been suggested as a prerequisite for survival of the causative organisms [22]. Still, the geographic distribution of the disease remains in large part uncharacterized. In this paper, we report explorations using ecological niche modeling to (1) estimate the current niche and potential distribution of mycetoma in an important endemic region (Sudan), (2) investigate risk factors associated with mycetoma infections in Sudan and South Sudan as reflected in distributional associations with environmental features, and (3) test *Acacia*-mycetoma associations based on overlap of the ecological niche of mycetoma infections with that of trees of the genus *Acacia*.

Materials and Methods

Occurrence records for mycetoma cases were obtained from published scientific literature via the PubMed database (www.ncbi.nlm.nih.gov/); we also used mycetoma data deposited in the GIDEON database (http://www.gideononline.com/). Studies were selected if they described positive mycetoma cases, and were referred to specific geographic locations that could be georeferenced precisely. When geographic references were textual in nature, latitude-longitude coordinates were assigned via reference to electronic gazetteers (e.g., http://www.fallingrain.com; [23]), and Google Earth (www.earth.google.com/); 11 records were obtained by georectification and georeferencing of Figure 1 from Ahmed *et al*. 2002 [4,17,24] . We eliminated duplicate records and records presenting obvious errors of identification prior to any further analysis.

Occurrence records were obtained for *Acacia* from the Global Biodiversity Information Facility (www.gbif.org) to test contributions of the trees to a robust mycetoma model for Sudan and South Sudan [4,17,24]. We filtered *Acacia* occurrences to include only Sudan and South Sudan. All duplicate records and records lacking georeferences were excluded from analysis.

To characterize environmental variation across Sudan and South Sudan, 8-day composite Land Surface Temperature and monthly Normalized Difference Vegetation Index (NDVI) data were drawn from Moderate Resolution Imaging Spectroradiometer (MODIS) satellite imagery at 1 km spatial resolution. We also used 10 variables from the World Soil Information site (http://www.isric.org) to summarize chemical and physical soil characteristics (S1 File). Soil data were obtained for each of 2 depths for each variable: 0-5 cm and 5-15 cm. Soil variables represented a collection of updatable soil property and class maps of the world at 1 km resolution

produced using model-based statistical methods, including 3D regressions with splines for continuous properties and multinomial logistic regression for classes [25].

LST and NDVI data were downloaded for 2005-2011 from the Land Processes Distributed Active Archive Center data holdings, using the NASA Reverb Echo data portal (https://reverb.echo.nasa.gov/reverb/) as described in greater detail elsewhere [26]. The LST product has been validated via several ground-truth and validation efforts over widely distributed locations and time periods [27]. The NDVI product has been used broadly for monitoring vegetation conditions and land cover change [28]. We calculated grids for the minimum, maximum, median, and ranges of values for LST and NDVI across the entire time sequence for all of Sudan and South Sudan to provide a rich characterization of environments across the country.

The Grinnellian fundamental ecological niche is defined by the set of coarse-grained, non-interactive environmental conditions under which a species is able to maintain populations without immigrational subsidy [29]. ENM attempts to estimate these niches from incomplete information by relating known occurrence locations and the environmental values that they present to the broader environmental landscape. This approach was used to relate known mycetoma occurrences to raster environmental data in an evolutionary-computing environment; in this case, a maximum entropy algorithm (MaxEnt v.3.3 [30]) was used to estimate ecological niches both for *Acacia* spp. collectively and for mycetoma. Niche model outputs for *Acacia* were in turn used as input in calibrating models for mycetoma; in the end, we developed models based on LST/NDVI and all combinations of soil and *Acacia* information, and the *Acacia* models were calibrated with and without soil information. Accessible areas (**M**) for mycetoma and *Acacia* were assumed to include all of Sudan and South Sudan, based on their wide geographic

distributions. We calibrated ENMs across a subset of the study region corresponding to known endemic districts; models were then transferred across all of Sudan and South Sudan for interpretation; for comparison, we also calibrated models across all of Sudan and South Sudan (i.e., not just known endemic districts), although the model transfer approach should be more rigorous [31]. ENMs outputs were converted to binary maps using a least training presence thresholding approach adjusted to admit 5% (E = 5%) error rates [32].

To test the ability of the ENM algorithm to predict occurrences accurately across unsampled areas of Sudan and South Sudan, we used a partial receiver operating characteristic (ROC) approach [32]. This approach evaluates models only over a range of relevant predictions, and potentially allows differential weighting of omission and commission errors, and therefore is preferable to traditional ROC approaches [32]. Models were evaluated by calibrating models with a random 50% of occurrences, and comparing the threshold-independent area under the curve (AUC) to null expectations. To compare partial ROC AUC ratios of each model with null expectations, the dataset was bootstrapped, and probabilities obtained by direct count, with AUC ratios calculated using a Visual Basic script developed by N. Barve (University of Kansas), based on 100 iterations and an E = 5% omission threshold.

As a further, and more rigorous, test of model predictivity, we derived a preliminary view of mycetoma case data archived in the Mycetoma Research Center, in Sudan, based on cases from 1991-2014. In view of the large scale of this data resource, we selected and georeferenced ~10% of the overall data archive at random; we eliminated cases lacking geographic references and removed records from duplicate localities, which left 158 localities for this preliminary analysis. We assessed the relationship of these data to the best of our model predictions via a one-tailed cumulative binomial probability calculation that assessed the probability of obtaining

the observed level of correct prediction by chance alone, given the background expectation of correct prediction based on the proportional coverage of the region by the prediction [29]..

Background similarity tests [33] were used to assess similarity between models of niches of *Acacia* and mycetoma. We first reassessed the accessible area (**M**) for both species [34]: mycetoma is limited approximately to the belt between the latitudes of 15^0 S and 30^0N [7,20], and *Acacia* is widely distributed and grows in a wide range of habitats [35], so we can set **M** as all of Sudan and South Sudan, or alternatively as only the known mycetoma-endemic districts (Fig 1). To test the null hypothesis of niche similarity between mycetoma and *Acacia* against the backgrounds of their respective **M** hypotheses [34] as described above, we used *D*-statistics and Hellinger's *I* implemented in ENMTools [33]. We tested niche similarity with respect to two environmental data sets: (1) LST and NDVI; and (2) LST, NDVI, and soil characteristics. The background similarity test is based on models of random points from across the accessible area in numbers equal to numbers of real occurrence data available for each species in the study, with 100 replicate samples. The null hypothesis of niche similarity was rejected if the observed *D* or *I* values fell below the 5^{th} percentile in the random-replicate distribution for comparison of the ENMs for the pair of species in question [33].

Results

We assembled a total of 44 records of mycetoma cases from sites across Sudan (Fig 1). Cases were from North Darfur (14), Gezira (8), North Kordufan (6), South Darfur (4), Sennar (3), and White Nile (3), Khartoum (2), River Nile (2), Kassala (1), and Northern (1) states. Sampling for mycetoma was focused in these regions, which can be considered as endemic districts for mycetoma. Mycetoma cases concentrated in a belt between 12°S and 19°N latitude, with only a

few cases outside this area in Sudan. Records for *Acacia* trees were obtained from 59 localities across Sudan and South Sudan (Fig 1). *Acacia* records were not limited to any particular sub-region, but rather were distributed across much of the country. The geographic distributions of *Acacia* trees and mycetoma cases appeared to overlap only in central Sudan. However, *Acacia* is also present in South Sudan, where no records were available for mycetoma.

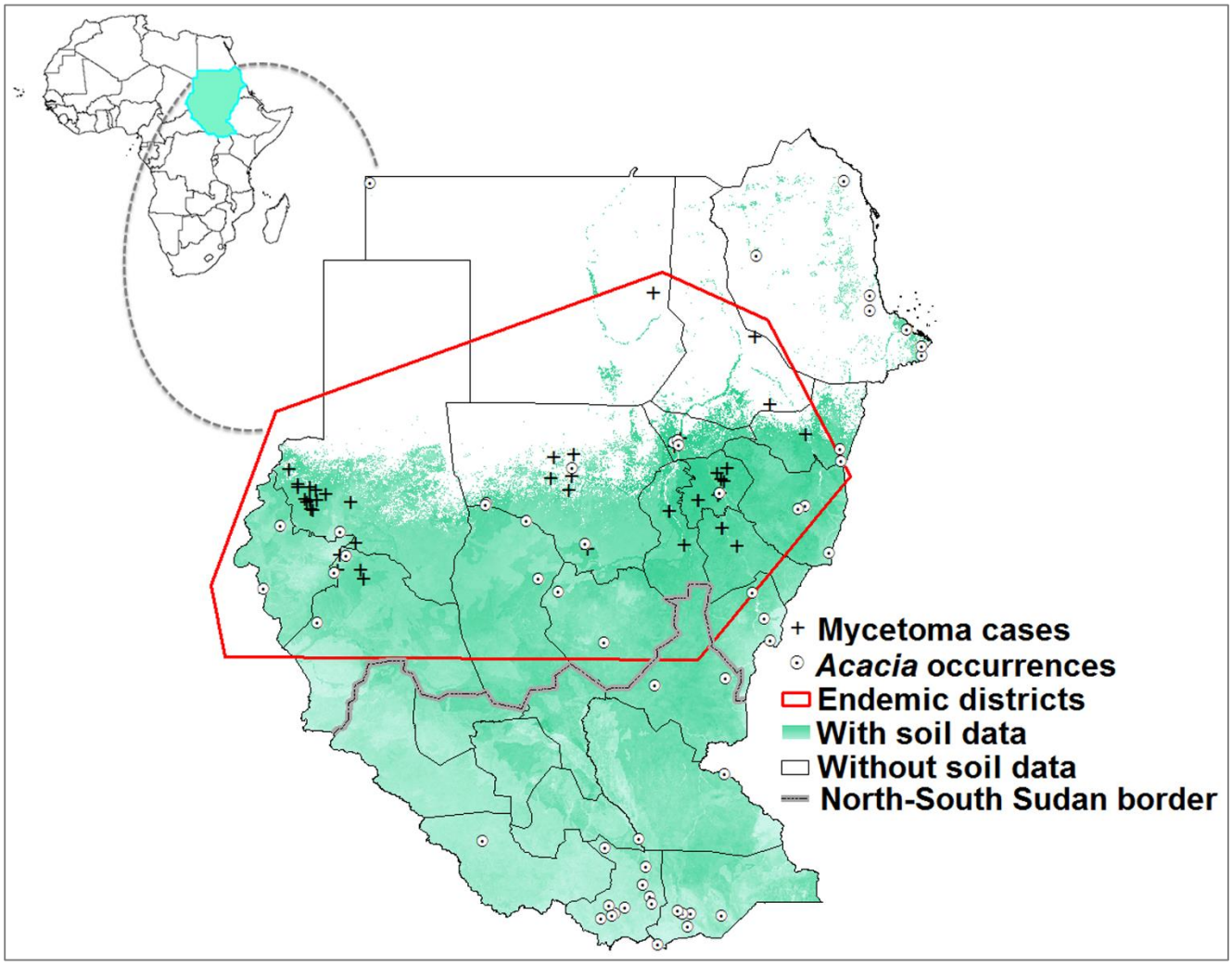


Fig 1: Geographic distribution of mycetoma cases and *Acacia* trees across Sudan and South Sudan (crosses and dotted circles, respectively). Some areas across the region (in white) were not included in some analyses for lack of data on soil characteristics.

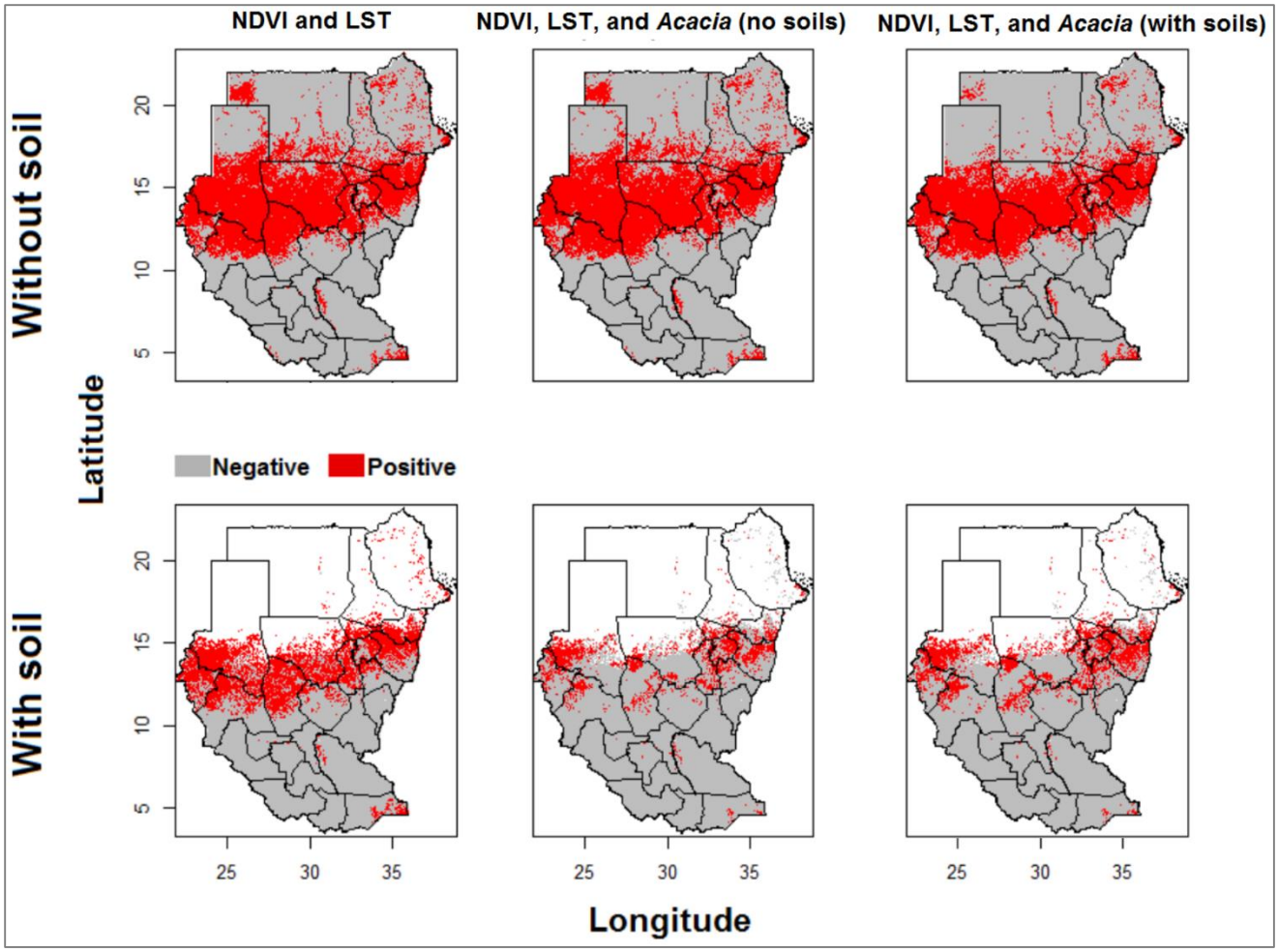


Fig 2: Potential mycetoma distribution based on occurrences in endemic districts. Potential distributions of mycetoma were based on different environmental variables; models were calibrated in mycetoma-endemic districts, and transferred across all of Sudan and South Sudan. White areas have no soils data, and therefore have no model predictions.

We developed models of mycetoma cases based on (1) ENMs calibrated in endemic districts, then transferred to all of Sudan and South Sudan (Fig 2), and (2) ENMs calibrated directly across all of Sudan and South Sudan; these latter models are not depicted in this chapter, but are presented in the supplementary materials (S2 File). ENMs for mycetoma based on different environmental scenarios were all statistically robust (all AUC ratios uniformly above 1.0 so all $P < 0.01$; see Table 1). The model based on all environmental data (LST, NDVI, soils, and *Acacia* distribution) had the highest partial AUC ratios, and thus appeared to perform best.

Mycetoma ENM predictions indicated a band of highest environmental suitability in central Sudan between 11°S and 17°N latitude (Figure 2). However, distinct areas were predicted as suitable for mycetoma occurrence elsewhere in Sudan and South Sudan: ENMs based on LST, NDVI, and soil identified a more southerly version of the "mycetoma belt." High-risk states identified by the ENMs included Kassala, Gedarif, Gezira, Khartoum, Sennar, White Nile, North Kordufan, West Kordufan, South Darfur, North Darfur, and West Darfur. To visualize ecological niches for mycetoma, we linked ENM predictions to characteristic of the environmental landscape (Fig 3): mycetoma occurs on diverse landscapes under wide ranges of environmental conditions, which is to say that no clear and distinctive environmental correlates could be discerned.

Environmental variables	AUC ratio
LST + NDVI	1.2923 (1.2917 – 1.8373)
LST + NDVI + soil	1.6864 (1.5179 – 1.9600)
LST + NDVI + *Acacia* (based on LST and NDVI)	1.3878 (1.3203 – 1.9011)
LST + NDVI + *Acacia* (based on LST, NDVI, and soil)	1.6518 (1.5767 – 1.8618)
LST + NDVI + soil + *Acacia* (based on LST and NDVI)	1.8402 (1.7683 – 1.9924)
LST + NDVI + soil + *Acacia* (based on LST, NDVI, and soil)	1.6365 (1.6012 – 1.8077)

Table 1: Partial AUC ratios of mycetoma ecological niche models based on different environmental data sets, showing median.

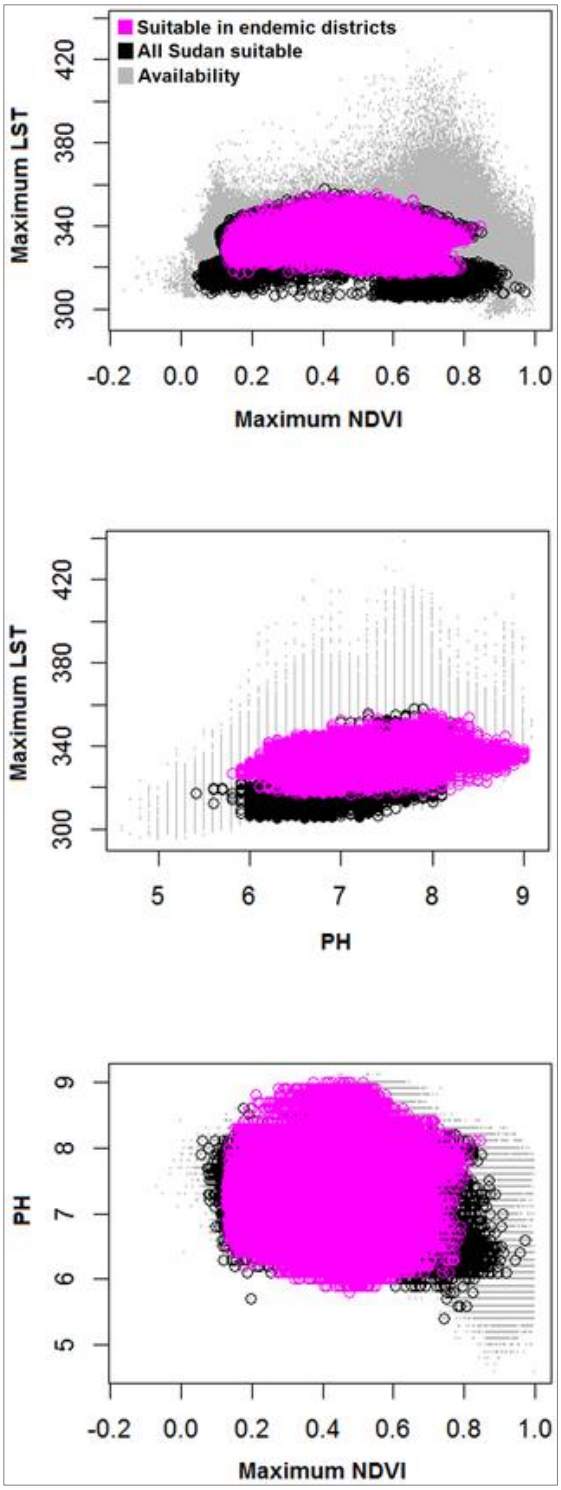


Fig 3: Visualization of mycetoma ecological niches (i.e., the set of environmental values under which the species can potentially maintain populations) in two-dimensional environmental spaces based on different environmental variables. The diagram shows the entire environmental availability across Sudan and South Sudan (light gray color), and conditions identified as suitable across Sudan and South Sudan (black color) and across endemic districts (pink).

Neither of the tests comparing niches of mycetoma and *Acacia* was able to reject the null hypothesis of niche similarity ($P > 0.05$ in both cases; Fig 4) which is to say that models for mycetoma and *Acacia* were not more different from one another than either was from models based on the background (i.e., across **M**) of the other species. *Acacia* is distributed broadly across Sudan and South Sudan, whereas mycetoma infections were found only in central Sudan, but these results suggest that range difference does not reflect niche differentiation between the two (sampling, diagnostic, and reporting biases may affect the mycetoma data).

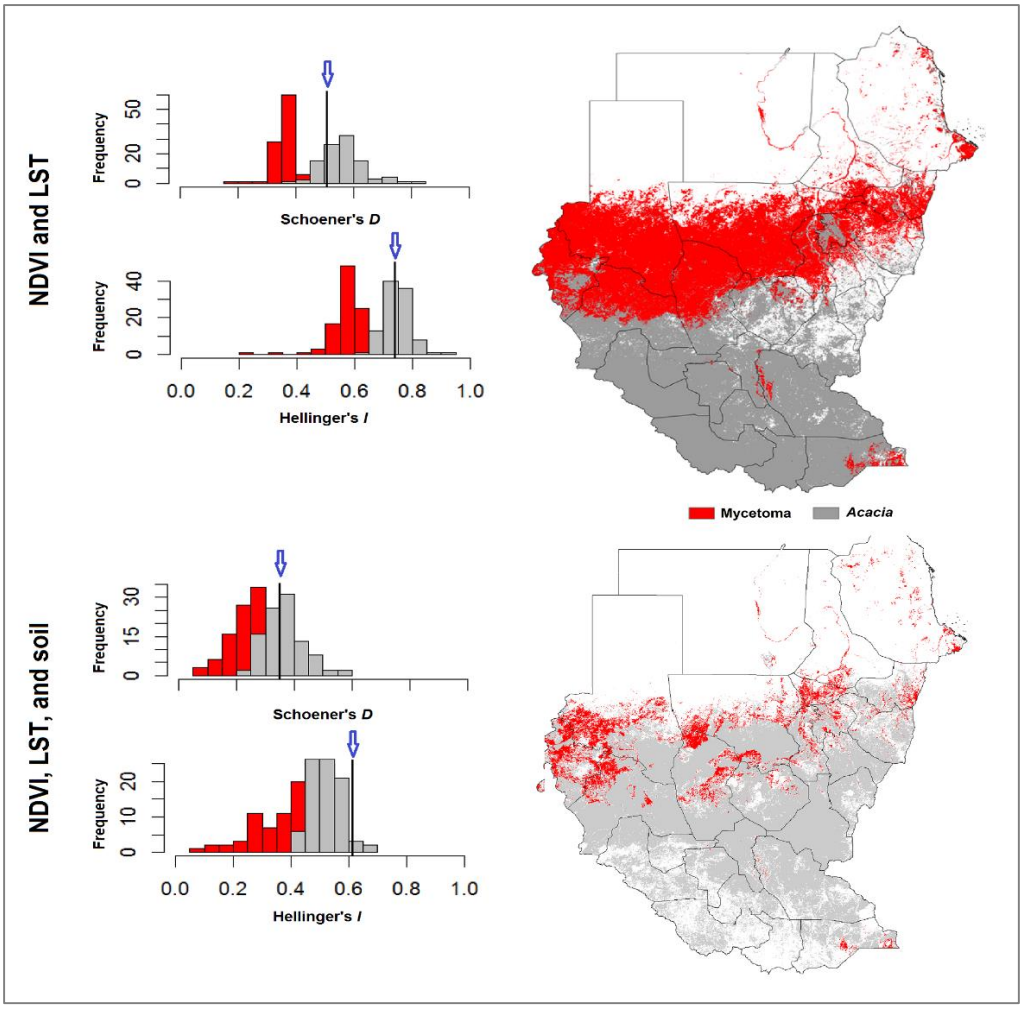


Fig 4: Background similarity test of similarity between mycetoma and *Acacia* ecological niches across Sudan and South Sudan. Niche overlap values were based on Hellinger's *I*, and Schoener's *D* metrics of similarity. Observed values are shown as black line with a blue arrow; null distribution is shown as a histogram.

The coincidence between model predictions and the independent additional case data from the Mycetoma Research Center was impressive (Fig 5), such that 149 of 158 of those additional occurrence points were successfully predicted by the model. Model success in anticipating these independent data was statistically significantly much better than random expectations (one-tailed cumulative binomial test; $P << 0.05$).

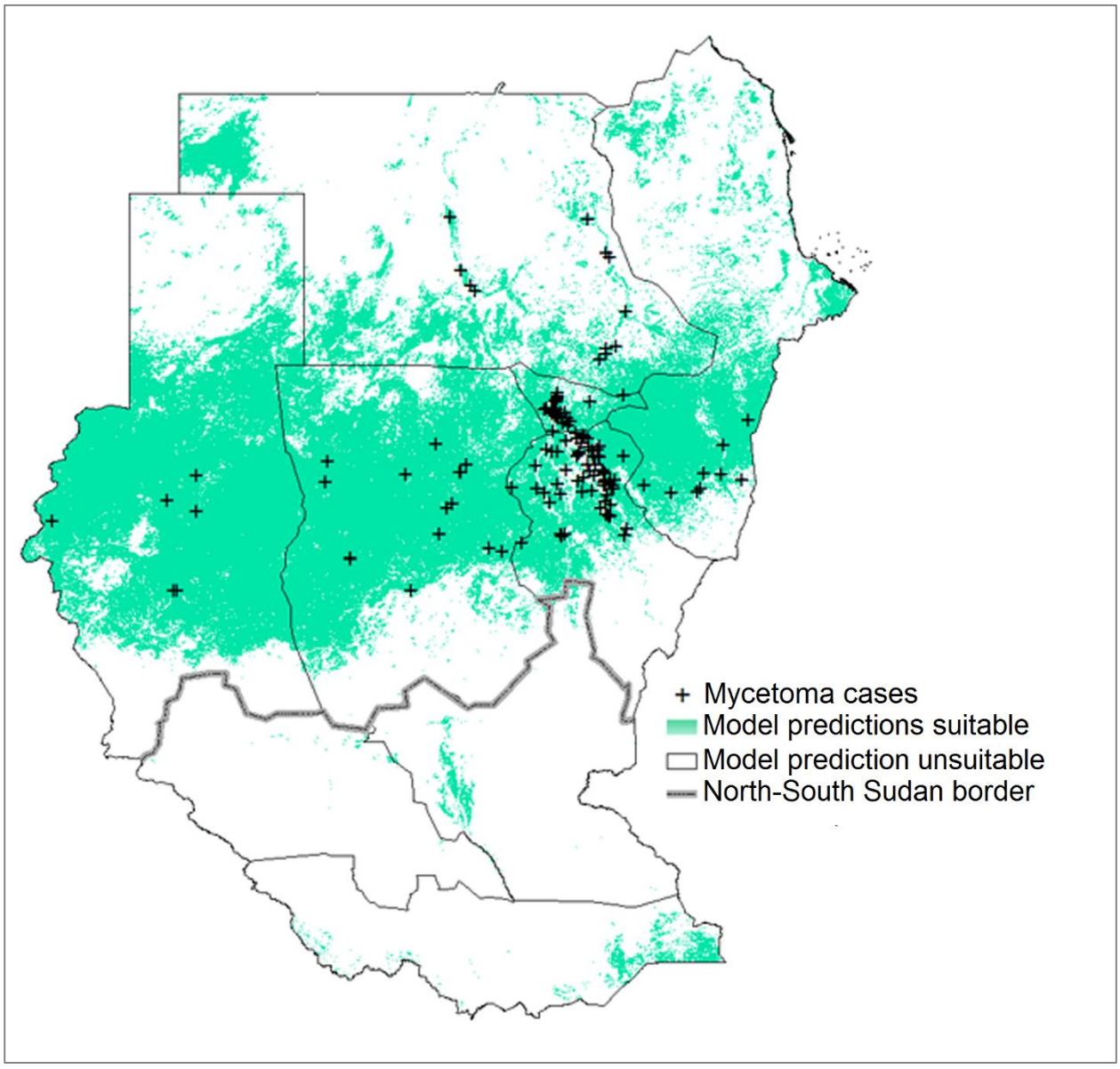


Fig 5: Coincidence between ecological niche model predictions based on LST, NDVI, soils, and Acacia (the latter based on LST and NDVI only) with the independent additional case data from the Mycetoma Research Center.

Discussion

Known since the 1600s [36] and described more formally in 1842, mycetoma was initially called Madura foot [37]. Mycetoma was subsequently reported in countries presenting diverse environments: Mexico, Venezuela, Mauritania, Senegal, Chad, Ethiopia, Sudan, Somalia, Yemen, and India [11,14]. Although thousands of cases have been recognized annually, risk factors remain poorly characterized [14], and the mode of transmission remains unknown [14]. Research on mycetoma leaves several hypotheses untested; improved understanding in each respect could reduce numbers of case, improve case outcomes, and offer possibilities for better disease control. Here, we used a new approach, termed ecological niche modeling, which relates case occurrences to environmental characteristics across a relevant region to create a model of the environmental 'envelope' (analogous to a coarse-grained definition of the ecological niche) for the species; this niche model allows, in turn, identification of potentially suitable areas for the species to be distributed. Ecological niche modeling has been used previously to understand geographic dimensions of a number of neglected tropical diseases [26,38,39], including fungal pathogens [40,41].

We used ENM to identify suitable sites for mycetoma infections based on environmental predictors, including dimensions thought to be associated with mycetoma cases in previous studies in Sudan [4,20]. All ENMs indicated high suitability across central Sudan, which appears consistent with cases reported subsequently [17,42,43]. It is worth noting that numerous cases reported by the Mycetoma Research Center (MRC) [4] came from the same belt identified by ENMs developed here, and yet had no involvement in our model calibration, providing important corroboration of the model predictions.

Several recent studies have attempted to understand modes of entry and transmission of mycetoma [4,44,45], but how people become infected with the causative agents remains unclear [14]. These studies have proposed that the primary reservoir of the causative agents is soil or *Acacia* thorns [4], and that transmission occurs by contact with the causative agent [4,15], based on observations that mycetoma infections occurred under poor conditions, in agriculturalists and villagers in endemic districts [46,47]. Our ENMs used soil data, but the causative agent has been identified from areas signaled unsuitable in the soil-based ENMs [4]. Incorporating *Acacia* distributions in models improved predictions, indicating possible relevance of an *Acacia*-mycetoma association.

Acacia may thus prove to play some role as a determinant of mycetoma distributional patterns across Sudan and South Sudan, although our results are correlational only and do not provide a direct test of this association. Our background similarity tests between ENMs for *Acacia* and mycetoma could not reject the hypothesis of similarity of the niches of the two species, thus at least not providing evidence against an association, and our models had greatest predictive power regarding mycetoma cases when *Acacia* distributions were included as environmental predictors. The important question remaining, however, is how the causative agent contacts humans, penetrates the skin, and initiates infections.

Previous studies confirmed presence of *Madurella mycetomatis* DNA in 17 of 74 soil samples and in one of 22 thorn samples [4]. Interestingly, attempts at culturing the fungi from these samples failed [4]. Hence, that the study found DNA of *M. mycetomatis* in both soil and thorn samples is of unclear importance, although perhaps culture methods are relatively insensitive or ineffective. In sum, then, our results revealed contributions of different environmental factors to mycetoma risk, identified areas suitable for mycetoma emergence,

farther raised the possibility of a mycetoma-*Acacia* association, and provided steps towards a robust predictive risk map for the disease.

Acknowledgments

The authors would like to thank the ENM Working Group in the Biodiversity Institute of the University of Kansas for their support and assistance during this work. Special thanks to the Department of Entomology and the Research and Training Center on Vectors of Diseases of Ain Shams University, Egypt. We also thank the staff of the Mycetoma Research Centre in Sudan.

References

1. Davis JD, Stone PA, McGarry JJ (1999) Recurrent mycetoma of the foot. J Foot Ankle Surg 38 (1): 55-60.

2. Pilsczek FH, Augenbraun M (2007) Mycetoma fungal infection: multiple organisms as colonizers or pathogens? Rev Soc Bras Med Trop 40 (4): 463-465.

3. Alam K, Maheshwari V, Bhargava S, Jain A, Fatima U, et al. (2009) Histological diagnosis of madura foot (mycetoma): a must for definitive treatment. J Glob Infect Dis 1 (1): 64-67.

4. Ahmed A, Adelmann D, Fahal A, Verbrugh H, van Belkum A, de Hoog S (2002) Environmental occurrence of *Madurella mycetomatis*, the major agent of human eumycetoma in Sudan. J Clin Microbiol 40 (3): 1031-1036.

5. Faqir F, Rahman Au (2004) Mycetoma: a local experience. J Postgrad Med Inst 18 (2): 172-175

6. Sahariah S, Sharma AK, Mittal VK, Yadav RV. (1978) Mycetoma of lower extremity. J Postgrad Med 24 (2): 113-116.

7. Lichon V, Khachemoune A (2006) Mycetoma: a review. Am J Clin Dermatol 7 (5):315-321.

8. Magana M (1984) Mycetoma. Int J Dermatol 23 (4): 221-236.

9. Brownell I, Pomeranz M, Ma L (2005) Eumycetoma. Dermatol Online J 11 (4): 10.

10. Negroni R, Lopez Daneri G, Arechavala A, Bianchi MH, Robles AM (2006) Clinical and microbiological study of mycetomas at the Muñiz Hospital of Buenos Aires between 1989 and 2004. Rev Argent Microbiol 38 (1): 13-18.

11. WHO (2013) The 17 neglected tropical diseases. Geneva: World Health Organization. Available: http://www.who.int/neglected_diseases/diseases/en/. Accessed 10 July 2014.

12. de Hoog GS, van Diepeningen AD, Mahgoub e-S, van de Sande WW (2012) New species of *Madurella*, causative agents of black-grain mycetoma. J Clin Microbiol 50: 988-994.

13. van de Sande WWJ (2013) Global burden of human mycetoma: a systematic review and meta-analysis. PLoS Negl Trop Dis 7: e2550.

14. van de Sande WWJ, Maghoub ES, Fahal AH, Goodfellow M, Welsh O, et al. (2014) The mycetoma knowledge gap: identification of research priorities. PLoS Negl Trop Dis 8: e2667.

15. Ezaldeen EA, Fahal AH, Osman A (2013) Mycetoma herbal treatment: the Mycetoma Research Centre, Sudan experience. PLoS Negl Trop Dis 7: e2400.

16. Fahal AH (2013) The Mycetoma Research Center, University of Khortum, Sudan: a success story that need support. Int J Sudan Res 3: 1-13.

17. Abd El-Bagi ME, Fahal AH (2009) Mycetoma revisited: incidence of various radiographic signs. Saudi Med J 30: 529-533.

18. de Hoog GS, Ahmed SA, Najafzadeh MJ, Sutton DA, Keisari MS, et al. (2013) Phylogenetic findings suggest possible new habitat and routes of infection of human eumyctoma. PLoS Negl Trop Dis 7: e2229.

19. Fahal AH, Hassan MA (1992) Mycetoma. Br J Surg 79 (11): 1138-1141.

20. Fahal AH (2004) Mycetoma: a thorn in the flesh. Trans R Soc Trop Med Hyg 98: 3-11.

21. van Belkum A, Fahal A, van de Sande WW (2013) Mycetoma caused by *Madurella mycetomatis*: a completely neglected medico-social dilemma. Adv Exp Med Biol 764: 179-189.

22. Ahmed AOA, van Leeuwen W, Fahal A, van de Sande W, Verbrugh H, et al. (2004) Mycetoma caused by *Madurella mycetomatis*: a neglected infectious burden. Lancet Infect Dis 4 (9): 566-574.

23. Wieczorek J, Guo Q, Hijmans R (2004) The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. Int J Geogr Inf Syst 18: 745-767.

24. Fahal A (2011) Mycetoma. Khartoum Med J 41: 514-523.

25. ISRIC-World Soil Information IWS (2013) Soil property maps of Africa at 1 km. Available for download at www.isric.org.

26. Samy AM, Campbell LP, Peterson AT (2014) Leishmaniasis transmission: distribution and coarse-resolution ecology of two vectors and two parasites in Egypt. Rev Soc Bras Med Trop 47: 57-62.

27. Coll C, Wan Z, Galve JM (2009) Temperature-based and radiance-based validations of the V5 MODIS land surface temperature product. J Geophys Res-Oc ATM 114: D20102.

28. Lyapustin AI, Wang Y, Laszlo I, Hilker T, G.Hall F, et al. (2012) Multi-angle implementation of atmospheric correction for MODIS (MAIAC): 3. Atmospheric correction. Remote Sens Environ 127: 385-393.

29. Peterson AT, Soberón J, Pearson RG, Anderson RP, Martínez-Meyer E, et al. (2011) Ecological Niches and Geographic Distributions. Princeton: Princeton University Press.

30. Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecol Model 190: 231-259.

31. Owens HL, Campbell LP, Dornak LL, Saupe EE, Barve N, et al. (2013) Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. Ecol Model 263: 10-18.

32. Peterson AT, Papeş M, Soberón J (2008) Rethinking receiver operating characteristic analysis applications in ecological niche modeling. Ecol Model 213: 63-72.

33. Warren DL, Glor RE, Turelli M (2008) Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. Evolution 62: 2868-2883.

34. Barve N, Barve V, Jiménez-Valverde A, Lira-Noriega A, Maher SP, et al. (2011) The crucial role of the accessible area in ecological niche modeling and species distribution modeling. Ecol Model 222: 1810-1819.

35. Aref IM, Atta H, Shahrani T, Mohamed A (2011) Effects of seed pretreatment and seed source on germination of five *Acacia* spp. Afr J Biotechnol 10: 15901-15910.

36. Kaempfer E (1694) Disputatio physica medica inauguralis exhibens decadem observationem exoticarum [phD thesis]. Netherlands: Univeristy of Leiden.

37. Gokhale BB (1981) Epidemiology of mycetoma. Hindustan Antibiot Bull 23: 18-24.

38. Peterson AT, Pereira RS, Neves VF (2004) Using epidemiological survey data to infer geographic distributions of leishmaniasis vector species. Rev Soc Bras Med Trop 37: 10-14.

39. Escobar LE, Peterson AT, Favi M, Yung V, Pons DJ, et al. (2013) Ecology and geography of transmission of two bat-borne rabies lineages in Chile. PLoS Negl Trop Dis 7: e2577.

40. Mak S, Klinkenberg B, Bartlett K, Fyfe M (2010) Ecological niche modeling of *Cryptococcus gattii* in British Columbia, Canada. Environ Health Persp 118: 653-658.

41. Reed KD, Meece JK, Archer JR, Peterson AT (2008) Ecologic niche modeling of *Blastomyces dermatitidis* in Wisconsin. PLoS ONE 3: e2034.

42. Ahmed AO, Desplaces N, Leonard P, Goldstein F, De Hoog S, et al. (2003) Molecular detection and identification of agents of eumycetoma: detailed report of two cases. J Clin Microbiol 41: 5813-5816.

43. Abd El-Bagi ME, Abdul Wahab O, Al-Thagafi MA, El-Sheikh H, Al-Salman K, Taifoor MK, Osman FM (2004) Mycetoma of the hand. Saudi Med J 25: 352-354.

44. Ahmed AO, van Vianen W, ten Kate MT, van de Sande WW, van Belkum A, et al. (2003) A murine model of *Madurella mycetomatis* eumycetoma. FEMS Immunol Med Microbiol 37 (1): 29-36.

45. Maiti PK, Bandyopadhyay D, Dey JB, Majumdar M (2003) Mycetoma caused by a new red grain mycetoma agent in two members of a family. J Postgrad Med 49: 322-324.

46. Chufal SS TN, Gupta MK (2012) An approach to histology-based diagnosis and treatment of Madura foot. J Infect Dev Ctries 6 (9): 684-688.

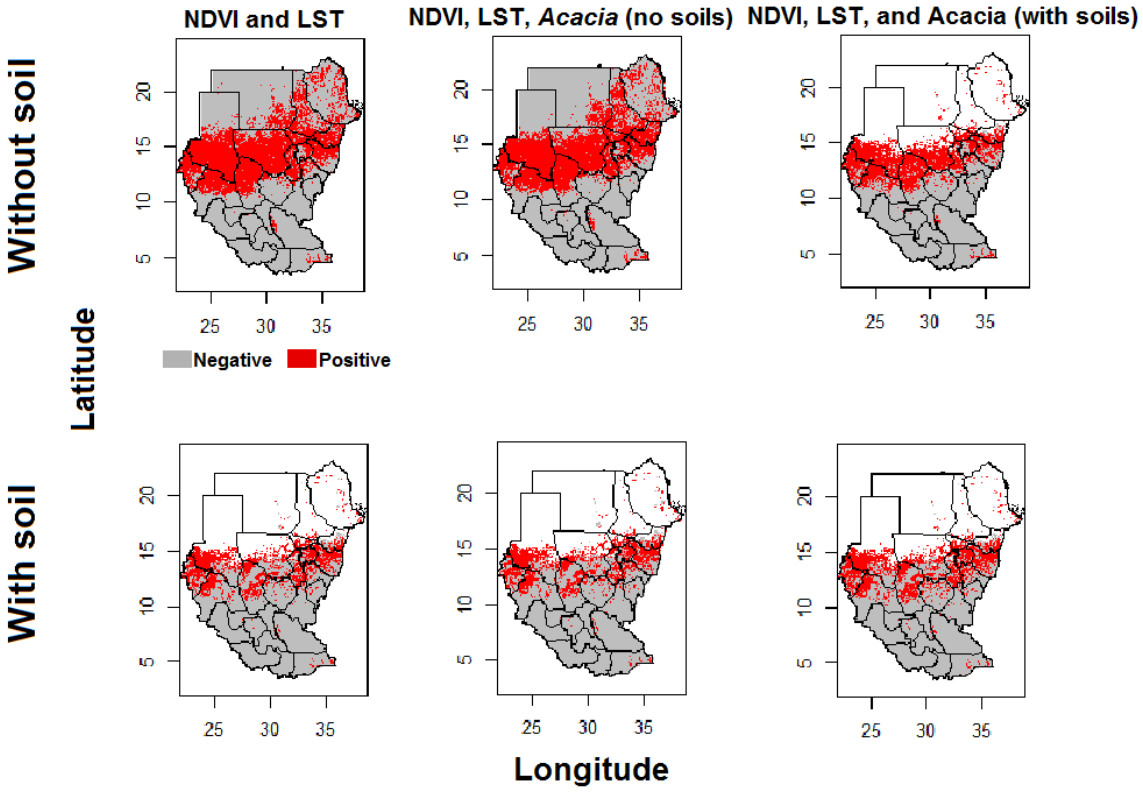
47. Maiti PK, Ray A, Bandyopadhyay S (2002) Epidemiological aspects of mycetoma from a retrospective study of 264 cases in West Bengal. Trop Med Int Health 7 (9): 788-792.

Supporting information

S1 File: The variables of the soil characteristics used in model calibration for mycetoma and *Acacia* spp. in Sudan. Data downloaded from the World Soil Information (http://www.isric.org). Each variable is available in 2 depths (0-5 cm and 5-15 cm).

Soil organic carbon in permilles (g/kg) = ORCDRC
predicted mean value for the first standard depth (0-5 cm)
predicted mean value for the first standard depth (5-15 cm)
pH in H2O 1:5 = PHIHO5
predicted mean value for the first standard depth (0-5 cm)
predicted mean value for the first standard depth (5-15 cm)
Sand content (50-2000 µm) in % = SNDPPT
predicted mean value for the first standard depth (0-5 cm)
predicted mean value for the first standard depth (5-15 cm)
Silt content (2-50 µm) in % = SLTPPT
predicted mean value for the first standard depth (0-5 cm)
predicted mean value for the first standard depth (5-15 cm)
Clay content (<2 µm) in % = CLYPPT
predicted mean value for the first standard depth (0-5 cm)
predicted mean value for the first standard depth (5-15 cm)

S2 File: Potential mycetoma distribution based on occurrences across all of Sudan. These models were calibrated across all of Sudan directly based on all records collected from scientific literature and environmental variables for all of Sudan.



Chapter 2: Coarse-resolution ecology of etiological agent, vector, and reservoirs of zoonotic cutaneous leishmaniasis in Libya[2]

[2] Samy AM, Annajar BB, Dokhan MR, Boussaa S, Peterson AT. (2016). Coarse-resolution Ecology of Etiological Agent, Vector, and Reservoirs of Zoonotic Cutaneous Leishmaniasis in Libya. PLoS Negl Trop Dis. 10(2):e0004381. doi: 10.1371/journal.pntd.0004381

Abstract

Cutaneous leishmaniasis ranks among the tropical diseases least known and most neglected in Libya. World Health Organization reports recognized associations of *Phlebotomus papatasi*, *Psammomys obesus*, and *Meriones* spp., with transmission of zoonotic cutaneous leishmaniasis (ZCL; caused by *Leishmania major*) across Libya. Here, we map risk of ZCL infection based on occurrence records of *L. major*, *P. papatasi*, and four potential animal reservoirs (*Meriones libycus*, *Meriones shawi*, *Psammomys obesus*, and *Gerbillus gerbillus*). Ecological niche models identified limited risk areas for ZCL across the northern coast of the country; most species associated with ZCL transmission were confined to this same region, but some had ranges extending to central Libya. All ENM predictions were significant based on partial ROC tests. As a further evaluation of *L. major* ENM predictions, we compared predictions with 98 additional independent records provided by the Libyan National Centre for Disease Control (NCDC); all of these records fell inside the belt predicted as suitable for ZCL. We tested ecological niche similarity among vector, parasite, and reservoir species and could not reject any null hypotheses of niche similarity. Finally, we tested among possible combinations of vector and reservoir that could predict all recent human ZCL cases reported by NCDC; only three combinations could anticipate the distribution of human cases across the country.

Introduction

Leishmaniasis remains one of the major public health problems in the Mediterranean Basin. In Libya, two forms of leishmaniasis occur: visceral leishmaniasis (VL), and cutaneous leishmaniasis (CL). VL has been reported in the country since 1904; however, little information is available on leishmaniasis epidemiology as regards the insect vector species and vertebrate reservoirs involved in transmission [1-3]. VL was identified from northeastern Libya and southern Saharan and sub-Saharan areas [1,4,5]. CL is most prevalent in the northwestern part of the country [2,6,7]. CL is caused by two species of *Leishmania*: *Leishmania major* Yakimoff & Schokhor, 1914 and *L. tropica* Wright, 1903 (Kinetoplastida: Trypanosomatidae). *Leishmania major* is the etiological agent of zoonotic CL (ZCL), where the parasite is thought to circulate in small-mammal reservoirs (*Meriones libycus* Lichtenstein, 1823 (Rodentia: Muridae), *Gerbillus gerbillus* Olivier, 1801 (Rodentia: Muridae), *Psammomys obesus* Cretzschmar, 1828 (Rodentia: Muridae), *M. shawi* Duvernoy, 1842 (Rodentia: Muridae)) and is transmitted by the sand fly *Phlebotomus papatasi* (Scopoli), 1786 (Diptera: Psychodidae) [2,7,8]. *Leishmania tropica* is the causative organism for anthroponotic CL (ACL); zoonotic foci have also been reported from rock hyrax in Kenya, and Israel [9,10], and gerbil in Egypt [11], where the disease is transmitted by the sand fly *P. sergenti* Parrot, 1917 (Diptera: Psychodidae) [10,11].

In Libya, seasonal wadis provide potential suitable conditions of climate and vegetation for vertebrate populations to maintain transmission [12]. The sand fly *P. papatasi* has a wide geographic distribution, from northern Africa to India [13]; it is considered as a proven vector of ZCL in North Africa [11]. In most field surveys in the country, *P. papatasi* and *P. sergenti* were the most abundant species [7,14,15]; however, *P. papatasi* was most frequent in the northern part of the country.

Recently, the World Health Organization identified four possible transmission systems of ZCL, based on associated mammal reservoirs: *Ps. obesus, Meriones* spp., *Rhombomys opimus* Lichtenstein, 1823 (Rodentia: Muridae), and *Mastomys* spp. Thomas, 1915 (Rodentia: Muridae) [16]. Limited epidemiological studies have been carried out in the country to characterize the roles of several species of reservoir hosts in maintaining CL in Libya. *Leishmania major* was identified from *Meriones libycus* [12], and *M. shawi* [16] in endemic areas of the northwestern part of the country. Early studies revealed *Ps. obesus* as the potential natural reservoir host of *L. major* in many North African countries including Libya [12,17]; *Ps. obesus* was most prevalent along wadi edges from Sahara to the Middle East, where high density of this species is associated with abundant vegetation and halophilic plants [17]. *Meriones* spp. are thought to play an important role in ZCL outbreaks by maintaining the parasite in nature in the long term. *Meriones shawi* and *M. libycus* have been found repeatedly to be naturally infected with *L. major* in Libya [12], Tunisia [18-20], Morocco [21], and Algeria [22].

Most ZCL outbreaks in North African countries have been tied to epidemiological modifications and environmental changes [23,24], highlighting the importance of understanding the epidemiology of ZCL in this region. This study represents a first effort to understand the ecology and geography of ZCL using remote-sensing data across Libya to predict ZCL risk areas. We used ecological niche modeling approaches to identify the distribution of sand fly vector species, mammal reservoirs, and the pathogens to test their patterns of overlap in environmental space, which illuminate details of the local ZCL cycle in Libya where these species coexist.

Materials and Methods

Study area

Libya is situated in North Africa on the Mediterranean coast between Egypt and Tunisia. The country lies between 18 and 33° N latitude and 8 and 25° E longitude. Dominant climate conditions include hot-summer Mediterranean and hot desert climates [25]: coastal lowlands have very hot summers and mild winters, while the desert interior has long, hot summers and high diurnal temperature ranges, with very dry conditions. Precipitation declines rapidly to the interior with distance from the coast. Libya lacks large rivers and streams, and extended droughts are frequent; however, the government has constructed a network of dams for water management [26].

Input data

Based on the leishmaniasis surveys in Libya, ZCL is endemic in the northwestern regions of the country. We collected records for all organisms involved in the ZCL transmission cycle including the pathogen (*L. major*), vector (*P. papatasi*), and potential mammal reservoirs (*Ps. obesus, M. libycus, M. shawi, G. gerbillus*). We retrieved vector and pathogen data from our own surveillance, and the PubMed database using keywords of species' names and Libya. When *L. major* was identified at the coarser district level (e.g. [2]), NCDC provided details for the exact locations of these cases for the purpose of this study. *Leishmania major* records based on clinical features only were excluded from analysis to avoid possible diagnostic errors in species identification; we included all records identified rigorously by either zymodeme analysis (i.e. MON-25) of 16 enzymatic loci [27] or restriction fragment length polymorphisms of the ribosomal internal transcribed spacer 1 (ITS1) region [2]. Host and vector species included in the

study were identified by reference to previously published morphological keys [28-30]. Data were included if a geographic reference was linked to any of the six species (geographic coordinates or textual descriptions). Other records of *Ps. obesus*, *M. libycus*, *M. shawi*, and *G. gerbillus* were obtained from the Global Biodiversity Information Facility (www.gbif.org), VertNet (http://www.vertnet.org/), and our own field surveillances across the country. When geographic references were textual in nature, we assigned longitude-latitude coordinates via reference to Google Earth (https://www.google.com/earth/). All occurrence data were filtered to eliminate duplicate records and longitude-latitude coordinates falling from outside Libya.

Environmental data sets by which to characterize environmental landscapes across Libya were obtained from three sources. (1) Advanced Very High Resolution Radiometer (NOAA-AVHRR) satellite imagery was obtained from the European Distributed Institute of Taxonomy (EDIT; http://bit.ly/1TDsUQM). These data comprise monthly mean Normalized Difference Vegetation Index (NDVI) coverage from 1982 to 2000, rescaled to a range of 1 to 255; we calculated mean, maximum, minimum, median, and range across the 12 monthly NDVI layers. (2) Climatic data layers representing 35 variables were obtained from global climatologies in CliMond (https://www.climond.org/; S1 File). (3) Digital elevation model were obtained from the Shuttle Radar Topography Mission (SRTM; http://srtm.usgs.gov/) at 1 km spatial resolution. All variables were resampled in ArcGIS 10.2 (Environmental Systems Resource Institute, Redlands, California) to a spatial resolution of 10 x 10' (≈20 x 20 km).

The particular environmental variables were chosen for modeling in light of their likely importance in shaping the geographic distributions of the species of interest in this study [25,31]. We selected historical NDVI and climatic data to cover the same time interval as when most records were obtained for the species. NDVI has been identified in previous epidemiological

studies as an important variable by which to convey seasonality resulting from changing temperature or moisture availability, and to understand broad-scale patterns of land use and land cover and their effects on pathogen populations and transmission [32]. NDVI is significantly correlated also with details of soil conditions, including type of soil, water content, and soil moisture [33-36]. Principal components analysis (PCA) was applied to the environmental variables to reduce multicollinearity and dimensionality. We used the first 10 principal components, which summarized more than 95% of the overall variance, to summarize environmental variation across Libya.

Ecological niche modeling

The MaxEnt algorithm [37] was used to estimate the fundamental ecological niche of the six species in this study. The fundamental ecological niche is defined as the set of environmental conditions under which a species is able to maintain populations without immigrational subsidy [38]. Correlational ecological niche models (ENMs) estimate niches by relating known occurrences to environmental values to identify conditions associated with the species presence. We calibrated ENMs within the districts where sampling was most detailed, and then transferred the model across all of Libya. MaxEnt was specified to conduct 100 bootstrapping replicates for each species. We used medians across the replicates as a final niche estimate for each species. All ENMs were converted to binary maps using a least training presence (i.e. lowest probability value of the occurrence points used in calibration of the models) thresholding approach adjusted to permit 5% omission in light of some probably erroneous records likely remaining in our data set [39].

Model Evaluation

To test the robustness of the ENMs in predicting the occurrences of the species accurately across unsampled areas of Libya, a partial receiver operating characteristic (ROC) approach was used [39]. This approach potentially allows differential weighting of omission (i.e., false negatives, leaving out actual distributional area) and commission errors (i.e., false positives, including unsuitable areas in prediction) and concentrates attention on parts of error space most relevant to niche modeling [39]. We selected 50% of the occurrence points of each species at random to test the ENMs by comparing the reduced threshold-independent area under the curve to null expectations: the dataset was bootstrapped, and probabilities obtained by direct count. AUC ratios were calculated via a software partial ROC available as a visual basic application at http://bit.ly/1JusDwz, based on 100 iterations and an E = 5% omission threshold.

An additional independent 98 records from the Libyan National Centre for Disease Control (NCDC) were used to test the model's ability to predict the distribution of new ZCL cases across Libya. These samples were identified in the NCDC laboratory based on PCR protocols from previous studies (e.g. [2,11]). We checked these records to remove any occurrences matching these used in calibrating ENMs, but none coincided with those used in model calibration. We used a one-tailed cumulative binomial probability distribution that assessed the probability of obtaining the observed level of correct prediction by a chance alone given the background expectation of correct predictions and based on the proportional coverage of the region by the thresholded model prediction.

Niche breadth and overlap

Niche breadth was estimated for each species based on the inverse concentration measure in ENMTools (http://enmtools.blogspot.com/)**.** For successful ZCL transmission, pathogen, vector,

and host species should overlap spatially and ecologically [31,40]. Here, ZCL transmission requires presence of *L. major*, *P. papatasi*, and at least one of the mammal reservoir species. We used background similarity tests [41] to assess similarity between pairs of estimated niches. We first estimated the accessible area (**M**) for each species in the study [42]; the accessible area for *L. major* was identified based on the distribution of that species across the country, where the species occurs only in the northwestern part. **M** estimates for the other species included all or at least a subset of the northern parts of the country depending on the species' current distributions. To test the null hypothesis of niche similarity between each pair of niches, we used *D*-statistics and Hellinger's *I* implemented in ENMTools [41]. Niche similarity was tested with respect to all environmental variables used to develop the ENM for each species. The background similarity test is based on generating random points from across the accessible area of one species in numbers equal to the numbers of real occurrence data available for that species in the study, with 100 replicate samples, and comparing an ENM based on these "background" points to the ENM of the other species. The null hypothesis of niche similarity was rejected if the *D* or *I* values fell below the 5^{th} percentile in the random-replicate distribution of similarity values [41].

We assumed that areas could be considered as at risk of ZCL transmission when all necessary elements for transmission co-occur [40]. We used the ENMs for *P. papatasi* and the four candidate ZCL reservoirs to identify areas of overlap between the vector and each of the possible hosts. These grids were obtained by multiplying the binary ENM of *P. papatasi* with the binary grid for the host species. We used a one-tailed cumulative binomial test to assess the relationship between the areas of vector-reservoir overlap and independent leishmaniasis human case records from the NCDC.

Results

We collected a total of 348 occurrences for *P. papatasi*, *L. major*, and four candidate reservoir species across Libya. Occurrence records are fully and openly available via Figshare repository (https://dx.doi.org/10.6084/m9.figshare.1613478). These data were concentrated along the northern coast of Libya (Fig 1). *Phlebotomus papatasi* was recorded from 84 localities, whereas *L. major* was characterized by 50 localities. *Meriones libycus* was the most commonly recorded mammal reservoir (104 sites) followed by *Ps. obesus* (48), *G. gerbillus* (32), and *M. shawi* (30). ENMs developed for these six species are illustrated in Fig 1; ENMs calibrated across the country (for comparison) are presented in the supplementary materials (S2 File).

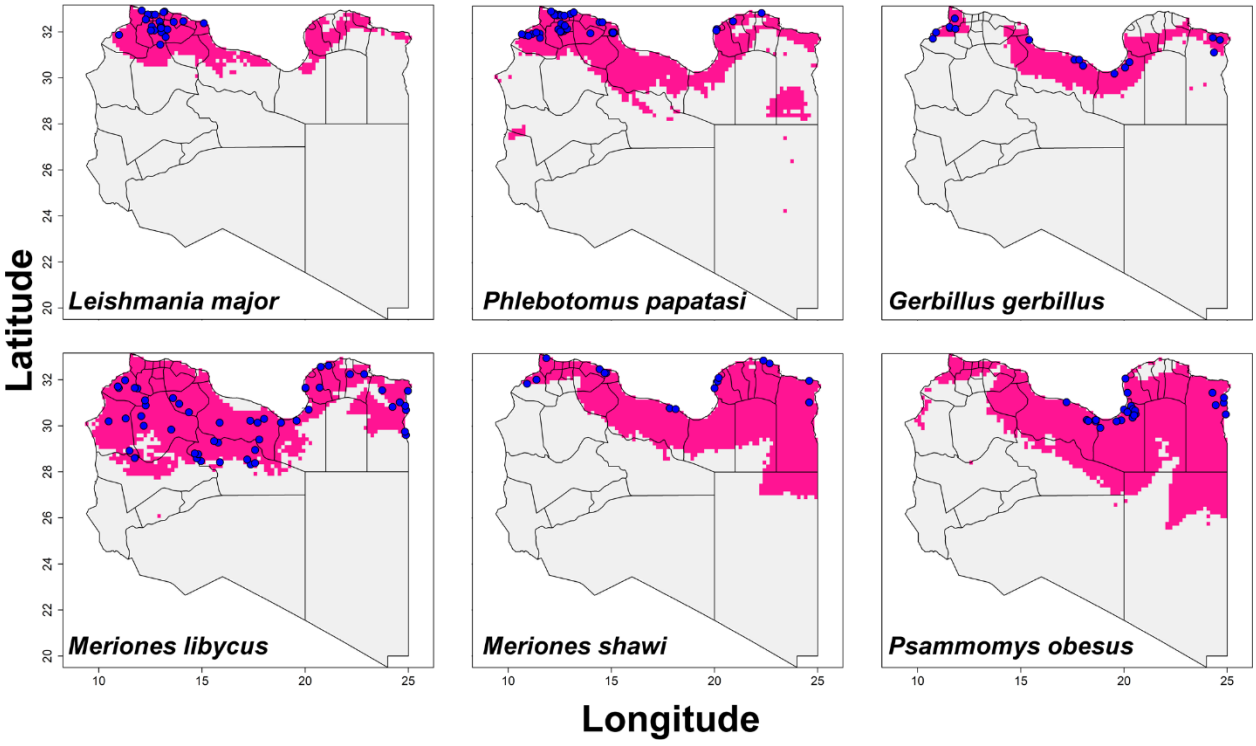


Fig 1: Thresholded potential distribution maps for *Leishmania major*, *Phlebotomus papatasi*, and four candidate mammal reservoir species potentially associated with the zoonotic transmission of cutaneous leishmaniasis. Models were calibrated across sampled area (S), and transferred across all Libya. Blue points are occurrences, pink areas are modeled suitable conditions, and gray areas are unsuitable conditions.

ENMs predicted most of the species to range across the northern coast of Libya; however, three species had broader potential distributions extending south to central Libya (*M. libycus*, *M. shawi*, *Ps. obesus*). The ENM for *L. major* predicted highest suitability in a belt between 30-33° N. These areas included many western provinces (e.g., Nalut, Yafran, Nuqat Al Khams, Al Jifarah, Sabratah, Misrata, Al Marqab, Gharyan, Babratah, Az Zawiya, Tajura, Tarhunati, Bani Walid, and Sirte), but also some eastern provinces (e.g., Al Jabal Al Akhdar, Al Qubbah, Al Hizam Al Akhdar, and Ajdabiya). The potential distribution of *P. papatasi* extended across the northern coast of Libya, but also in a disjunct area in central east Libya. All ENMs calibrated for these species were significantly robust based on partial ROC tests, with AUC ratios uniformly above 1 ($P < 0.001$; Table 1).

Species	Mean	Minimum	Maximum
Leishmania major	1.92	1.91	1.95
Phlebotomus papatasi	1.94	1.93	1.98
Gerbillus gerbillus	1.57	1.50	1.88
Meriones libycus	1.56	1.35	1.89
Psammomys obesus	1.71	1.64	1.96
Meriones shawi	1.78	1.72	1.94

Table 1: Results of partial ROC analysis to test statistical significance of ecological niche model predictions. A value of 1.0 is equivalent to the performance of a random classifier. These results were based on 100 bootstrap replicates, and statistical significance was assessed via bootstrapping and comparison with a random classifier ratio of 1.0.

In the most recent CL outbreaks across Libya, NCDC identified *L. major* in cases from 98 sites. The *L. major* ENM predicted 98 out of 98 of these additional independent data, which is statistically better than random expectations ($P < 0.001$). These additional independent data thus corroborated the *L. major* ENM, and the ability of that model to anticipate all recent cases of ZCL identified (Fig 2).

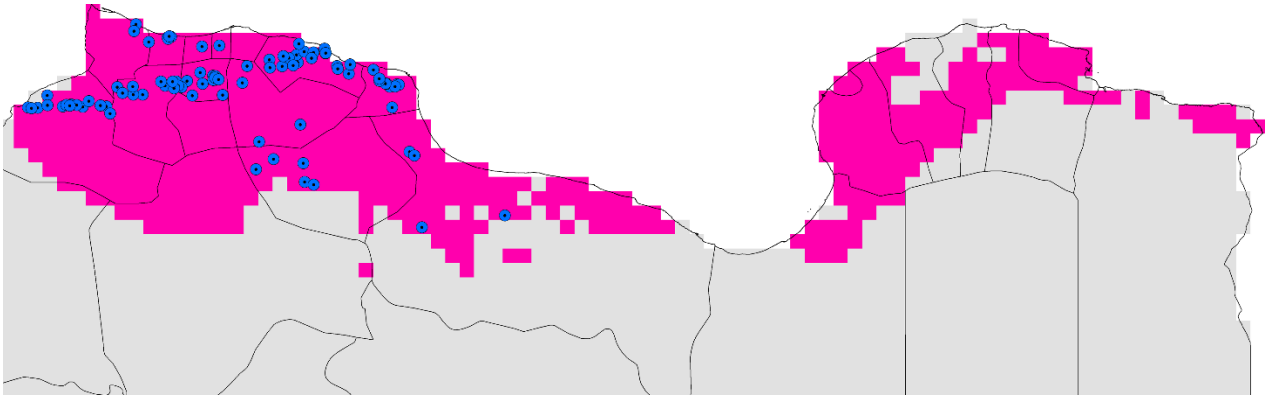


Fig 2: Relationship of ecological niche modeling predictions to the distribution of 98 sites with *L. major* cases reported by the Libyan National Centre for Disease Control in recent outbreaks across Libya. The blue dotted circle represented localities where these independent data were collected, and pink represent the belt predicted suitable for the *Leishmania major*.

Niche breadth was least in *L. major* and *P. papatasi*, and greater in the mammal species; indeed only *G. gerbillus* had niche breadth similar to *L. major* (S3 File). We visualized the environmental conditions where these species occur: *L. major* and *P. papatasi* were at low elevations, and mostly under a maximum temperature of 25 °C – 37 °C (Fig 3). The other species had similar responses to environmental conditions; however, they tend to be distributed along a broader environmental range (except *G. gerbillus*; S4 File).

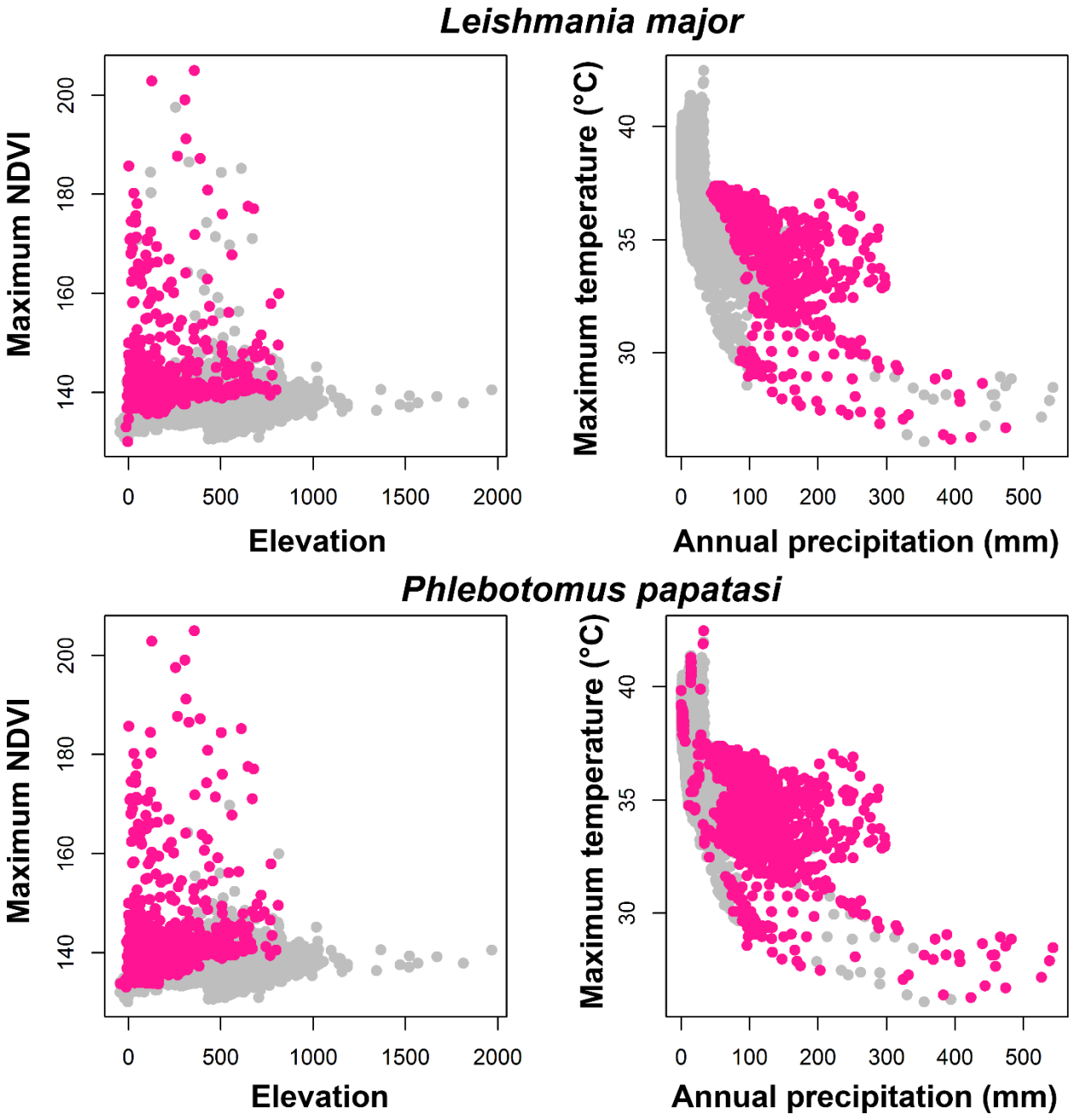


Fig 3: Visualization of *Leishmania major*, and *Phlebotomus papatasi* ecological niches in example dimensions. Overall set of environments available across Libya in gray; modeled suitable conditions for the species occurrences in pink. Similar visualizations of ecological niches for the potential mammal reservoir species are in the Supporting Information (S4 File).

The background similarity tests comparing the ENMs of parasite, vector, and possible reservoirs were uniformly unable to reject the null hypothesis of niche similarity between these species ($P > 0.05$; Fig 4 & S5 File). This result indicates the niche estimate for *L. major* could not be distinguished from those of the vector or the four potential reservoirs. We used NicheA to visualize overall overlap between the species based on three dimensions of PCAs (Fig 5), which revealed broad overlap in environmental conditions used by six species.

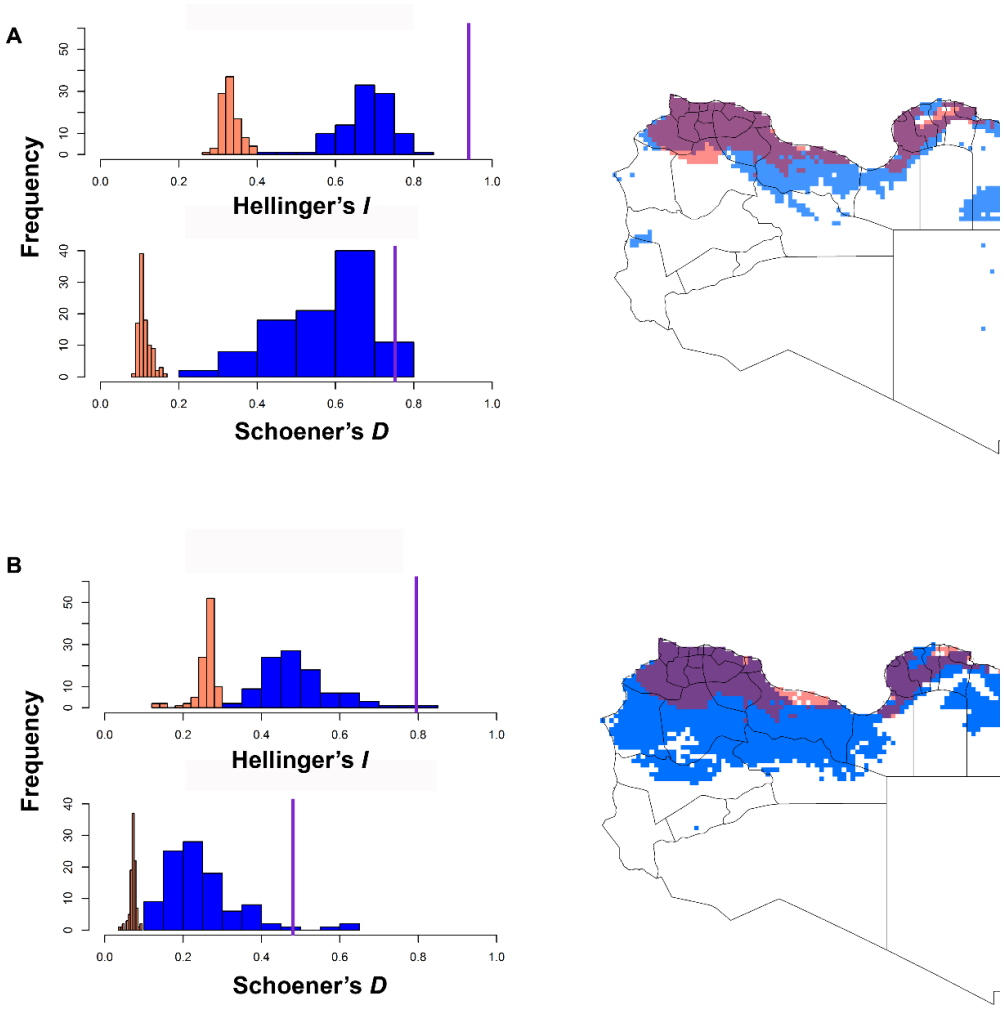


Fig 4: Example background similarity tests showing overall niche overlap between ecological niche models for pairs of species: (A) *Leishmania major—Phlebotomus papatasi* and (B) *Leishmania major –Meriones libycus*. The vertical purple line shows observed niche overlap, and the histograms show the distribution of the background similarity values among 100 random replicates, for the *I* and *D* similarity metrics. On the maps, red and blue shading indicates the modeled suitable areas for the two species; purple shading shows areas of overlap between the two species. Results for other species are given in the Supporting Information (S5 File).

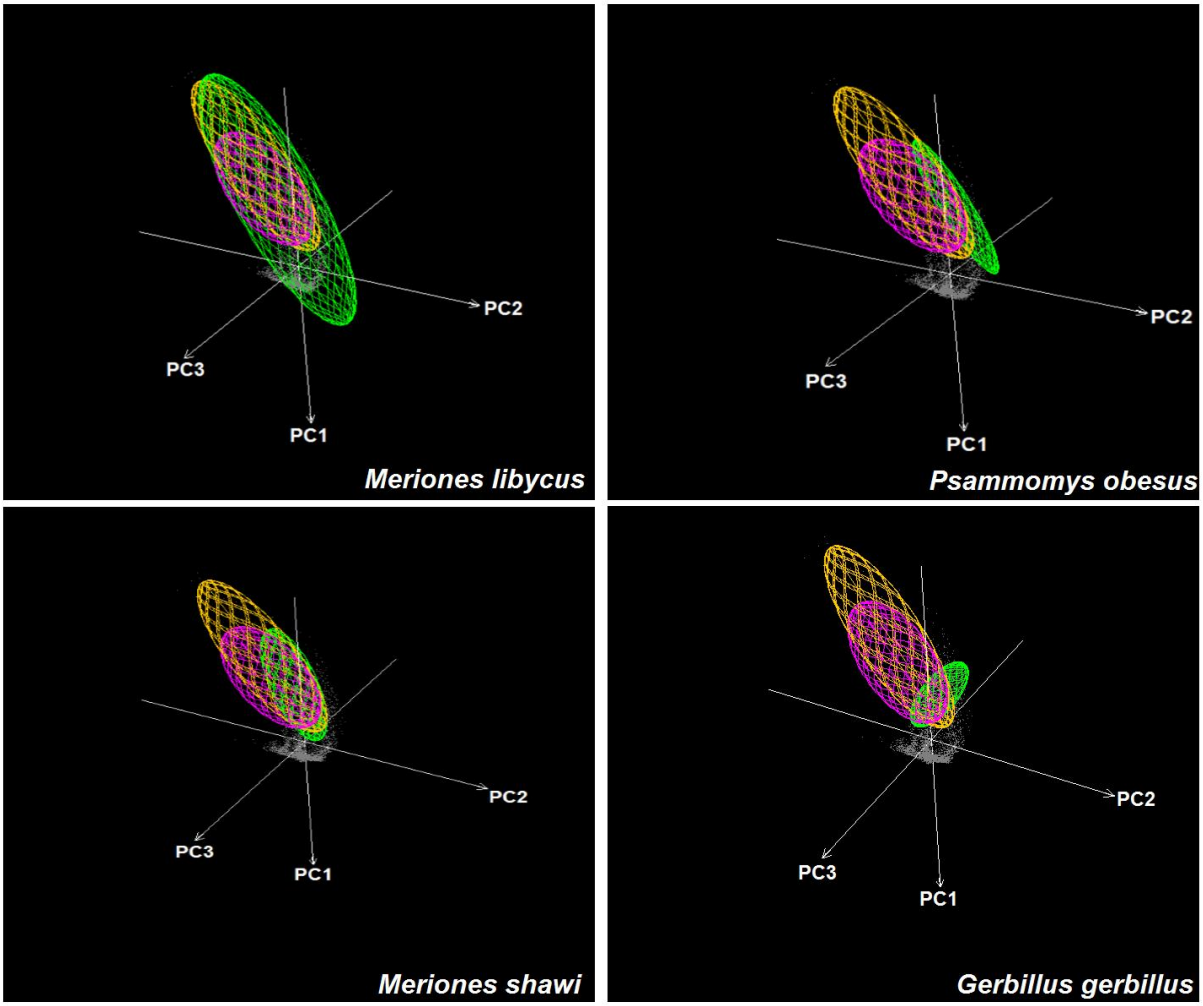


Fig 5: Visualization of ecological niches of *Leishmania major*, *Phlebotomus papatasi*, and animal reservoir in three environmental dimensions (PC1, PC2, and PC3). Niches are represented as minimum volume ellipsoids to illustrate the limits under which the species has been sampled. Gray shading represents environmental background, green ellipsoid represents the potential mammal reservoir, yellow is the vector *Phlebotomus papatasi*, and purple represents *Leishmania major*.

Finally, we combined the modeled distribution of the vector *P. papatasi* with those of each of the potential reservoirs as hypotheses of system that could support zoonotic transmission of CL across Libya (Fig 6). Results revealed that *P. papatasi-M. libycus*, *P. papatasi-M. shawi*, and *P. papatasi-Ps. obesus* systems predicted recent ZCL well ($P < 0.01$); the first two predicted 100% of the cases reported to the NCDC, but *Ps. obesus* identified only 85.7% of these cases (84

out of 98). The *P. papatasi-G. gerbillus* map was able to predict only 29 of 98 records, not better than null expectations ($P > 0.05$).

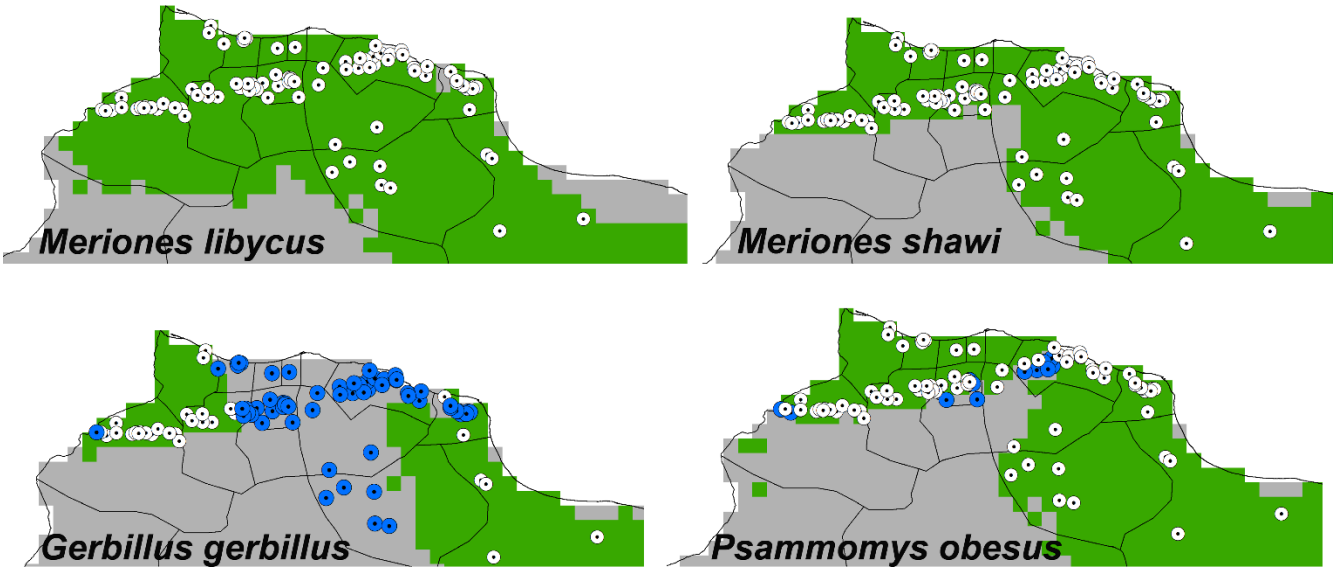


Fig 6: Relationship of additional independent human case records to the areas where pairs of vector *Phlebotomus papatasi* and mammal reservoir species can occur. Green areas are areas of overlap between *P. papatasi* and each of the potential mammal reservoirs; white dotted circle represent localities where human cases were predicted successfully; blue dotted circles indicate case records not predicted successfully by the model combination.

Discussion

Numerous recent studies have attempted to map potential distributions of key species involved in leishmaniasis transmission in several countries in Europe and the Americas [43,44]. Africa, however, has seen only a few efforts to map vector populations [31,45-47]. Libya sees many CL cases [3]; for example, 6284 cases were identified there in 2006 alone (S6 File). CL case rates are still underestimated owing to inefficient infrastructure for early notifications of cases, and lack of public awareness among doctors and patients [3]. For the national control program to be successful, all organisms associated with leishmaniasis transmission should be identified and understood in detail (i.e. vectors, reservoirs, and pathogens).

We developed this mapping exercise across Libya for several reasons. (1) Most prominently, we wished to map the potential distribution of ZCL cases across the country. (2) We strove to map the potential distribution of 5 other organisms potentially associated with the disease's dynamics in Libya. (3) We wished to test niche similarity among the set of species involved. Finally, (4) we tested the possible reservoir-vector combinations that could allow better prediction of ZCL cases. All of these analyses will help to understand the disease risk areas across the country and guide possible control programs.

Our models identified risk areas across both the western and eastern portions of the north coast of the country. Although all previous studies in Libya had found CL cases only in the western provinces, some recent reports have provided evidence of CL occurrence in eastern sites as well (e.g. Ajdabiya, and Al Jabal Al Akhdar; [48]). Although this report [48] is the only one to place CL at these sites, most CL surveillance has concentrated in western Libya [2,6,7], so this results is perhaps expected. Our ENMs found suitability of both regions for ZCL transmission, benefitting from higher-resolution environmental data, and consideration of areas that were sampled and accessible to each species [42,49].

The risk of ZCL transmission in North Africa appears to be determined by the joint dynamics of vectors and mammal reservoir populations [16]. When we visualized the environmental conditions suitable for the species examined in this study, they were most prevalent in a maximum temperature range of 25 °C – 37 °C, similar to other recent reports across North Africa and Middle East [50-52]. These latter studies reported that *P. papatasi* was abundant in semi-arid and arid steppe zones, and that low and high temperatures are key in limiting its distribution and activity [50-52]. For example, *P. papatasi* in Morocco was less active at temperatures of 11–20 °C and 37–40 °C [52]. The distributional patterns of *L. major* and *P.*

papatasi estimated in this study concords with these latter reports [51,52]. Northern coastal regions of Libya are characterized by a Mediterranean climate, whereas the rest of the country has hot, dry desert climates that are unfavorable for these species, with maximum summer temperatures over 40 °C apparently. Previous studies have shown that water is a major limiting factor for sand flies and for leismaniasis abundance and spread, respectively [53]. *Phlebotomus papatasi* cannot tolerate the extreme conditions of temperature and low humidity associated with the rare rainfall in the south, although the species is well established in other deserts where conditions are more mesic (e.g., Negev Desert [53]). Our study identified an interesting prediction of the presence of suitable environmental conditions in central Libya, associated with construction of new water resources [26] and raised concerns for changes in the eco-epidemiology of leishmaniasis across the country as water resources (S7 File) and agricultural activities are established in southern parts of the country. These important anthropogenic changes will be key factors in affecting distributions of vectors and reservoir hosts of ZCL across Libya; for example, in other studies in the region, soil moisture was an important variable in determining vector and reservoir abundance [54,55]; anthropogenic disturbance was also identified as favoring conditions for vector and larger host populations in Israel [56]. The effects of these two factors may be reflected among some of environmental variables included in our study, but their absence in explicit terms still marks a limitation to our study; a more detailed picture of ZCL transmission risk in the region will need to consider their possible effects on long-term sand fly and rodent abundances.

Most recent ZCL cases occurred at relatively low elevations; the areas near Al Jabal Al Gharbi alone accounted for most cases (S7 File) [2,57]. Similar observations were reported for *L. major*, *P. papatasi*, and wild mammals in Morocco [58]. Elevation and temperature are not the only factors influencing the distribution of ZCL cases: precipitation has also been shown to play

a role [45]. Low-elevation northern areas, where *L. major* and *P. papatasi* species occur in high densities, are characterized by the highest precipitation in the country [45].

Although testing niche similarity among species was unable to distinguish among hypotheses of ZCL hosts, as was possible in our previous analyses in Egypt [31], our analyses of possible species combinations excluded *G. gerbillus* as a main reservoir across Libya. In our previous analysis in Egypt, however, we found marked niche similarity between *P. papatasi* and *L. major*, but none between *L. major* and *P. sergenti* in terms of geographic distribution and ecological niche [31], supporting the idea that carefully constructed ENMs are able to predict disease risk based on models of vectors and reservoir hosts in a complex transmission system like leishmaniasis. In this study, 85.7% of cases were predicted successfully, focusing on areas where *Ps. obesus* co-occurred with the vector. WHO had reported that *Ps. obesus* was likely the main reservoir of ZCL in Libya [16]; however, our results more strongly supported the two *Meriones* spp. – *P. papatasi* system, which were able to anticipate all recent human cases. Evidence for this association has also been found in the form of high infection rates with *L. major* in *Meriones tristrami* in the most recent ZCL foci documented in Israel [59]; these observations provide mounting evidence that jird play a major role in disease transmission across the region. This study took Libya as a target population for illuminating the identity and distribution of reservoir hosts in the complex ZCL cycle. Indeed, simply the definition of "reservoir host" remains unresolved [60-62]; early studies defined reservoir host as the "ecological system in which the infection agent survives indefinitely" [60], but later studies focused on definition in reference to a specific target population [61]. Certainly, some confusions in reservoir definition still exist; the specification of particular target populations emphasizes the importance of geographic and ecological associations in defining reservoir hosts, which

underlines the approach in this study. As a result, we urge development of similar studies regarding other target populations to examine spatial and temporal relationships of these hosts, and characterize differences in ZCL dynamics among regions.

The study of the association among these organisms in both spatial and temporal dimensions is of great added values to map the ZCL risk areas across Libya, guide the control program across the country, and provide the first detailed maps for the potential distributions of organisms associated with the zoonotic transmission cycle across Libya. An early study shed light on disease ecology and possible host-pathogen associations [63], discussing criteria of host geographic distribution, pathogen range within the host range, regional distributions of organisms in different biomes and habitats, relative prevalence of the pathogen among host subpopulations, temporal and fine-scale spatial pattern of host-pathogen dynamics, and integrative time- and place-specific predictive models. These criteria were discussed as major steps to promote understanding of pathogen-host associations in complex transmission cycles. This study applied most of these criteria to the complex ZCL cycle in Libya but we note knowledge gaps in Libya regarding the prevalence of *L. major* among different host subpopulations, and the dynamics and potential distribution of host and parasite at finer scales across the country. Filling these gaps as regards the disease system in Libya will promote a more detailed picture both for its ecology and for control programs.

Leishmaniasis control programs should consider our findings by applying integrated approaches to combating ZCL by considering the environmental risk factors that we have explored. That is, if a particular combination of host and vector species is necessary for leishmaniasis transmission, then strategies by which to interrupt that transmission can focus on removing the pathogen, the vector, or key hosts from the system. Such measures may be

implemented via educational programs in risk areas, mass drug administration in infected communities, and host or vector control programs. Our future work will focus on possible hotspots in the less-well-known areas of the country via intensive disease surveillance and sampling of all relevant organisms. More deeply, we plan to consider socioeconomic variables in tandem with the physical environmental variables for a more universal model that links physical, biological, and human factors in this complex disease system.

Acknowledgments

The authors would like to thank the KU ENM group, and the Department of Entomology and the Research and Training Center on Vectors of Diseases of Ain Shams University for their continuous support during this work. Special thanks go also to all staff members of the Libyan National Centre for Disease Control for their support.

References

1. Mehabresh MI (1994) Visceral leishmaniasis: new foci of infection in Libya. J Trop Med Hyg 97: 282-285.
2. Amro A, Gashout A, Al-Dwibe H, Zahangir Alam M, Annajar B, et al. (2012) First molecular epidemiological study of cutaneous leishmaniasis in Libya. PLoS Negl Trop Dis 6 (6): e1700.
3. Alvar J, Velez ID, Bern C, Herrero M, Desjeux P, et al. (2012) Leishmaniasis worldwide and global estimates of its incidence. PLoS One 7 (5): e35671.

4. Jain S, el Mangoush MA, el-Bouri K, Mahfouz MO (1990) Kala azar in an adult Libyan and review of visceral leishmaniasis in Libya. Trop Geogr Med 42: 283-285.

5. Mehabresh MI, el-Mauhoub MM (1992) Visceral leishmaniasis in Libya--review of 21 cases. Ann Trop Paediatr 12: 159-163.

6. Ashford RW, Chance ML, Ebert F, Schnur LF, Bushwereb AK, et al. (1976) Cutaneous leishmaniasis in the Libyan Arab Republic: distribution of the disease and identity of the parasite. Ann Trop Med Parasitol 70: 401-409.

7. el-Buni AA, Jabeal I, Ben-Darif AT (2000) Cutaneous leishmaniasis in the Libyan Arab Jamahiriya: a study of the Yafran area. East Mediterr Health J 6: 884-887.

8. Belal US, Abdel-Hafeez EH, Naoi K, Norose K (2012) Cutaneous leishmaniasis in the Nalut District, Libyan Arab Jamahiriya: a clinico-epidemiologic study and Leishmania species identification. J Parasitol 98: 1251-1256.

9. Sang DK, Njeru WK, Ashford RW (1992) A possible animal reservoir for *Leishmania tropica* s.l. in Kenya. Ann Trop Med Parasitol 86: 311-312.

10. Jacobson RL (2003) *Leishmania tropica* (Kinetoplastida: Trypanosomatidae)--a perplexing parasite. Folia Parasitol 50: 241-250.

11. Shehata MG, Samy AM, Doha SA, Fahmy AR, Kaldas RM, et al. (2009) First report of *Leishmania tropica* from a classical focus of *L. major* in North-Sinai, Egypt. Am J Trop Med Hyg 81: 213-218.

12. Ashford RW, Schnur LF, Chance ML, Samaan SA, Ahmed HN (1977) Cutaneous leishmaniasis in the Libyan Arab Republic: preliminary ecological findings. Ann Trop Med Parasitol 71: 265-271.

13. Maroli M, Feliciangeli MD, Bichaud L, Charrel RN, Gradoni L (2013) Phlebotomine sandflies and the spreading of leishmaniases and other diseases of public health concern. Med Vet Entomol 27: 123-147.

14. Obenauer PJ, Annajar BB, Hanafi HA, Abdel-Dayem MS, El-Hossary SS, et al. (2012) Efficacy of light and nonlighted carbon dioxide-baited traps for adult sand fly (Diptera: Psychodidae) surveillance in three counties of Mesrata, Libya. J Am Mosq Control Assoc 28: 179-183.

15. Ahmed S, Abou faddan H (2013) Cutaneous leishmaniasis in Gharyan – Libya – a case-control study. Life Sci 10: 826-834.

16. WHO (2010) Control of the leishmaniases. Report of a meeting of the WHO Expert Committee on the Control of Leishmaniases. Available at http://bit.ly/1FqYJTI17.

17. Aoun K, Bouratbine A (2014) Cutaneous leishmaniasis in North Africa: a review. Parasite 21: 14.

18. Rioux JA, Petter F, Zahaf A, Lanotte G, Houin R, et al. (1986) Isolation of Leishmania major Yakimoff and Shokhor, 1914 (Kinetoplastida-Trypanosomatidae) in *Meriones shawi-shawi* (Duvernoy, 1842) (Rodentia-Gerbillidae) in Tunisia. Ann Parasitol Hum Comp 61: 139-145.

19. Ben-Ismail R, Helal H, Kouzena N, Ben Rachid MS (1987) Natural infestation of *Meriones libycus* in a focus of cutaneous zoonotic leishmaniasis in Douara (Tunisia). Ann Soc Belg Med Trop 67: 201-202.

20. Ghawar W, Toumi A, Snoussi MA, Chlif S, Zaatour A, et al. (2011) *Leishmania major* infection among *Psammomys obesus* and *Meriones shawi*: reservoirs of zoonotic

cutaneous leishmaniasis in Sidi Bouzid (central Tunisia). Vector Borne Zoonotic Dis 11: 1561-1568.

21. Rioux JA, Petter F, Akalay O, Lanotte G, Ouazzani A, et al. (1982) *Meriones shawi* (Duvernoy, 1842) [Rodentia, Gerbillidae] a reservoir of *Leishmania major*, Yakimoff and Schokhor, 1914 [Kinetoplastida, Trypanosomatidae] in South Morocco (author's transl). C R Seances Acad Sci III 294: 515-517.

22. Boudrissa A, Cherif K, Kherrachi I, Benbetka S, Bouiba L, et al. (2012) Spread of *Leishmania major* to the north of Algeria. Bull Soc Pathol Exot 105: 30-35.

23. Bounoua L, Kahime K, Houti L, Blakey T, Ebi KL, et al. (2013) Linking climate to incidence of zoonotic cutaneous leishmaniasis (*L. major*) in pre-Saharan North Africa. Int J Environ Res Public Health 10: 3172-3191.

24. Lotfy WM (2014) Climate change and epidemiology of human parasitosis in Egypt: A review. J Adv Res 5: 607-613.

25. Kottek M, Grieser J, Beck C, Rudolf B, Rubel F (2006) World Map of the Köppen-Geiger climate classification updated. Meteorol Z 15: 259-263.

26. Mgherbi OK (2000) Rehabilitation and reconstruction of the Wadi Gattara dams in Libya. Int J Hydro Dams 7: 67-69.

27. Annajar B (1999) Epidemiology of cutaneous leishmaniasis in Libya. PhD dissertation, Keele University, Staffordshire, United Kingdom.

28. Lewis DJ (1982) A taxonomic review of the genus *Phlebotomus* (Diptera: Psychodidae) Bull Brit Mus (Nat Hist), Entomol Ser 45.121-209

29. Lane RP (1986) The sand flies of Egypt (Diptera: Phlebotominae). Bull Brit Mus (Nat Hist), Entomol Ser 52: 1-35

30. Osborn DJ, Helmy I (1980) Mammals of Egypt. Chicago: Field Museum of Natural History: 1–579

31. Samy AM, Campbell LP, Peterson AT (2014) Leishmaniasis transmission: distribution and coarse-resolution ecology of two vectors and two parasites in Egypt. Rev Soc Bras Med Trop 47: 57-62.

32. Peterson AT (2013) Mapping Disease Transmission Risk: Enriching Models Using Biogeography and Ecology Johns Hopkins University Press, Baltimore, Maryland, USA, 2014.

33. Todd SW, Hoffer RM (1998) Responses of spectral Indices to variations in vegetation cover and soil Background. Photogramm Eng Remote Sens 64: 915-921.

34. Scanlon TM, Albertson JD, Caylor KK, Williams CA (2002) Determining land surface fractional cover from NDVI and rainfall time series for a savanna ecosystem. Remote Sens Environ 82: 376-388.

35. McNally A, Funk C, G. J. Husak GJ, Michaelsen J, B. Cappelaere, et al. (2013) Estimating Sahelian and East African soil moisture using the Normalized Difference Vegetation Index. Hydrol Earth Syst Sci Discuss 10: 7963–7997.

36. Zhang L, Ji L, Wylie BK (2011) Response of spectral vegetation indices to soil moisture in grasslands and shrublands. Int J Remote Sens 32: 5267-5286.

37. Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecol Model 190: 231-259.

38. Peterson AT, Soberón J, Pearson RG, Anderson RP, Martínez-Meyer E, et al. (2011) Ecological Niches and Geographic Distributions; Levin SA, Horn HS, editors. Princeton and Oxford: Princeton University.

39. Peterson AT, Papeş M, Soberón J (2008) Rethinking receiver operating characteristic analysis applications in ecological niche modeling. Ecol Model 213: 63-72.

40. Peterson AT (2008) Biogeography of diseases: a framework for analysis. Naturwissenschaften 95: 483-491.

41. Warren DL, Glor RE, Turelli M (2008) Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. Evolution 62: 2868-2883.

42. Barve N, Barve V, Jiménez-Valverde A, Lira-Noriega A, Maher SP, et al. (2011) The crucial role of the accessible area in ecological niche modeling and species distribution modeling. Ecol Model 222: 1810-1819.

43. Chamaille L, Tran A, Meunier A, Bourdoiseau G, Ready P, et al. (2010) Environmental risk mapping of canine leishmaniasis in France. Parasit Vectors 3: 31.

44. Machado da Silva AV, Magalhaes Mde A, Pecanha Brazil R, Carreira JC (2011) Ecological study and risk mapping of leishmaniasis in an endemic area of Brazil based on a geographical information systems approach. Geospat Health 6: 33-40.

45. Abdel-Dayem MS, Annajar BB, Hanafi HA, Obenauer PJ (2012) The potential distribution of *Phlebotomus papatasi* (Diptera: Psychodidae) in Libya based on ecological niche model. J Med Entomol 49: 739-745.

46. Thomson MC, Elnaiem DA, Ashford RW, Connor SJ (1999) Towards a kala azar risk map for Sudan: mapping the potential distribution of *Phlebotomus orientalis* using digital data of environmental variables. Trop Med Int Health 4: 105-113.

47. Elnaiem DE, Schorscher J, Bendall A, Obsomer V, Osman ME, et al. (2003) Risk mapping of visceral leishmaniasis: the role of local variation in rainfall and altitude on the presence and incidence of kala-azar in eastern Sudan. Am J Trop Med Hyg 68: 10-17.

48. Pigott DM, Golding N, Messina JP, Battle KE, Duda KA, et al. (2014) Global database of leishmaniasis occurrence locations, 1960-2012. Sci Data 1.

49. Owens HL, Campbell LP, Dornak LL, Saupe EE, Barve N, et al. (2013) Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. Ecol Model 263: 10-18.

50. Killick-Kendrick R (1999) The biology and control of phlebotomine sand flies. Clin Dermatol 17: 279-289.

51. Wasserberg G, Yarom I, Warburg A (2003) Seasonal abundance patterns of the sandfly *Phlebotomus papatasi* in climatically distinct foci of cutaneous leishmaniasis in Israeli deserts. Med Vet Entomol 17: 452-456.

52. Boussaa S, Guernaoui S, Pesson B, Boumezzough A (2005) Seasonal fluctuations of phlebotomine sand fly populations (Diptera: Psychodidae) in the urban area of Marrakech, Morocco. Acta Trop 95: 86-91.

53. Schlein Y, Warburg A, Schnur LF, Le Blancq SM, Gunders AE (1984) Leishmaniasis in Israel: reservoir hosts, sandfly vectors and leishmanial strains in the Negev, Central Arava and along the Dead Sea. Trans R Soc Trop Med Hyg 78: 480-484.

54. Fichet-Calvet E, Jomâa I, Zaafouri B, Ashford RW, Ben-Ismail R, et al. (2000) The spatio-temporal distribution of a rodent reservoir host of cutaneous leishmaniasis. J Appl Ecol 37: 603-615.

55. Wasserberg G, Abramsky Z, Anders G, El-Fari M, Schoenian G, et al. (2002) The ecology of cutaneous leishmaniasis in Nizzana, Israel: infection patterns in the reservoir host, and epidemiological implications. Int J Parasitol 32: 133-143.

56. Wasserberg G, Abramsky Z, Kotler BP, Ostfeld RS, Yarom I, et al. (2003) Anthropogenic Disturbances Enhance Occurrence of Cutaneous Leishmaniasis in Israel Deserts: Patterns and Mechanisms. Ecol Appl 13: 868-881.

57. Global Infectious Diseases and Epidemiology Network (GIDEON) (2014). Available at http://www.gideononline.com/.

58. Guernaoui S, Boumezzough A, Laamrani A (2006) Altitudinal structuring of sand flies (Diptera: Psychodidae) in the High-Atlas mountains (Morocco) and its relation to the risk of leishmaniasis transmission. Acta Trop 97: 346-351.

59. Faiman R, Abbasi I, Jaffe C, Motro Y, Nasereddin A, et al. (2013) A newly emerged cutaneous leishmaniasis focus in northern Israel and two new reservoir hosts of *Leishmania major*. PLoS Negl Trop Dis 7(2):e2058.

60. Ashford RW (1997) What it takes to be a reservoir host. Belg J Zool 127: 85–90

61. Haydon DT, Cleaveland S, Taylor LH, Laurenson MK (2002) Identifying reservoirs of infection: a conceptual and practical challenge. Emerg Infect Dis 8: 1468-1473.

62. Ashford RW (2003) When is a reservoir not a reservoir?: Emerg Infect Dis. 9(11). 1495-6.

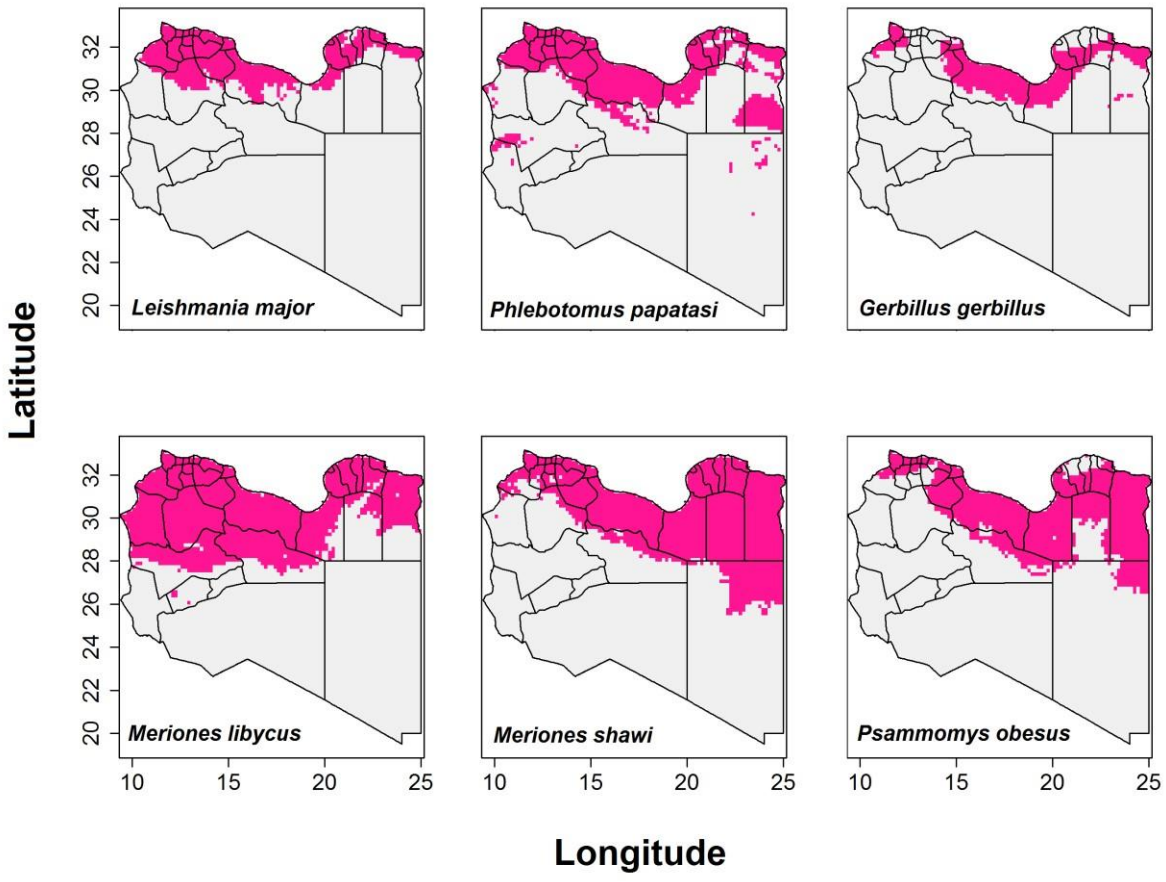
63. Mills JN, Childs JE (1998) Ecologic studies of rodent reservoirs: their relevance for human health. Emerg Infect Dis 4: 529-537.

Supporting information

S1 File: Detailed description of the CliMond variables used in the model. Details on these variables are also available via https://www.climond.org/.

Variable Number	Variable
Bio01	Annual mean temperature (°C)
Bio02	Mean diurnal temperature range (mean(period max-min)) (°C)
Bio03	Isothermality (Bio02 ÷ Bio07)
Bio04	Temperature seasonality (C of V)
Bio05	Max temperature of warmest week (°C)
Bio06	Min temperature of coldest week (°C)
Bio07	Temperature annual range (Bio05-Bio06) (°C)
Bio08	Mean temperature of wettest quarter (°C)
Bio09	Mean temperature of driest quarter (°C)
Bio10	Mean temperature of warmest quarter (°C)
Bio11	Mean temperature of coldest quarter (°C)
Bio12	Annual precipitation (mm)
Bio13	Precipitation of wettest week (mm)
Bio14	Precipitation of driest week (mm)
Bio15	Precipitation seasonality (C of V)
Bio16	Precipitation of wettest quarter (mm)
Bio17	Precipitation of driest quarter (mm)
Bio18	Precipitation of warmest quarter (mm)
Bio19	Precipitation of coldest quarter (mm)
Bio20	Annual mean radiation (W m^{-2})
Bio21	Highest weekly radiation (W m^{-2})
Bio22	Lowest weekly radiation (W m^{-2}
Bio23	Radiation seasonality (C of V)
Bio24	Radiation of wettest quarter (W m^{-2})
Bio25	Radiation of driest quarter (W m^{-2})
Bio26	Radiation of warmest quarter (W m^{-2})
Bio27	Radiation of coldest quarter (W m^{-2})
Bio28	Annual mean moisture index
Bio29	Highest weekly moisture index
Bio30	Lowest weekly moisture index
Bio31	Moisture index seasonality (C of V)
Bio32	Mean moisture index of wettest quarter
Bio33	Mean moisture index of driest quarter
Bio34	Mean moisture index of warmest quarter
Bio35	Mean moisture index of coldest quarter

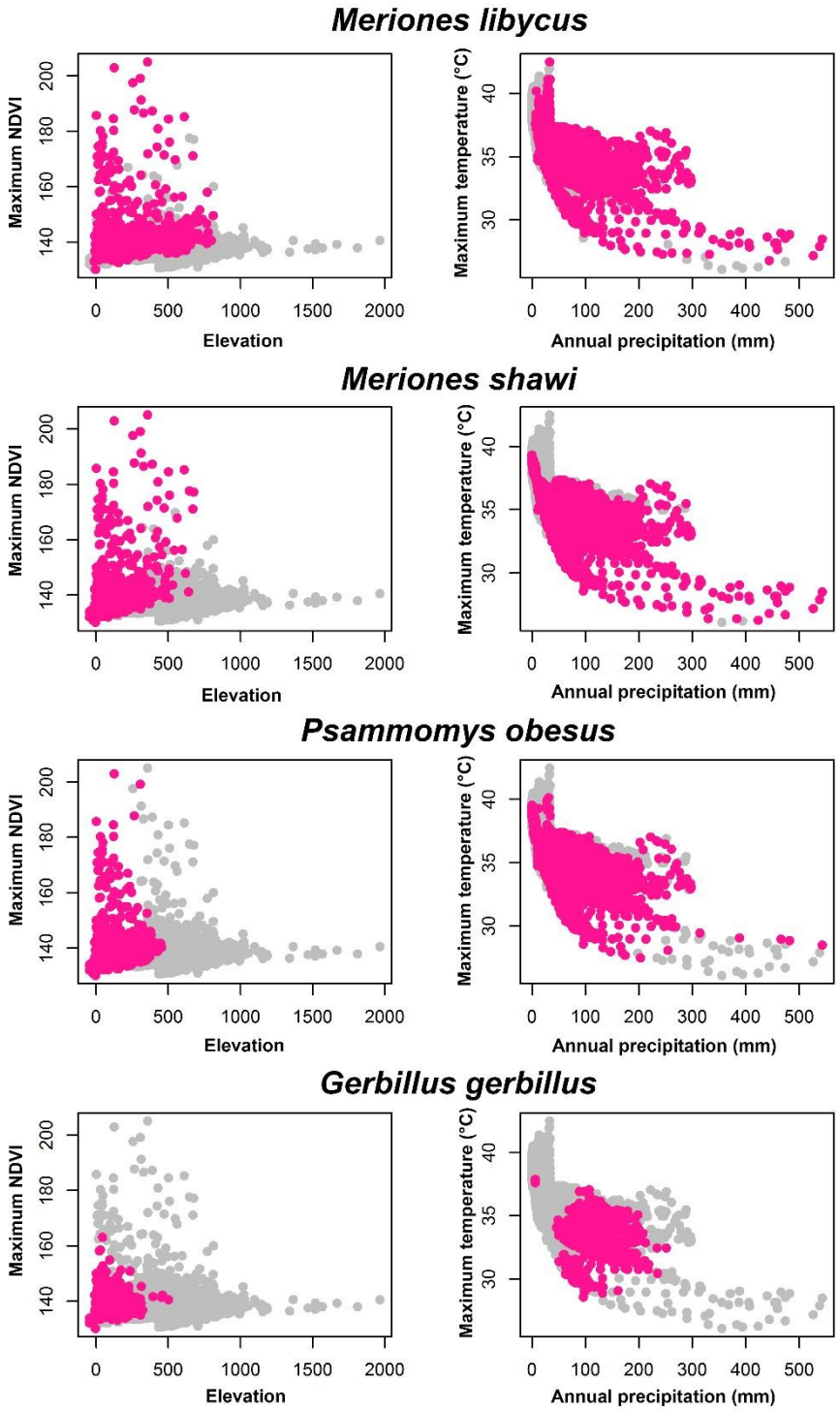
S2 File: Thresholded potential distribution maps for *Leishmania major*, *Phlebotomus papatasi*, and four candidate mammal reservoir species potentially associated with the zoonotic transmission of cutaneous leishmaniasis. Models were calibrated directly across Libya. The pink areas represent modeled suitable conditions, and gray areas were modeled as unsuitable for the species.



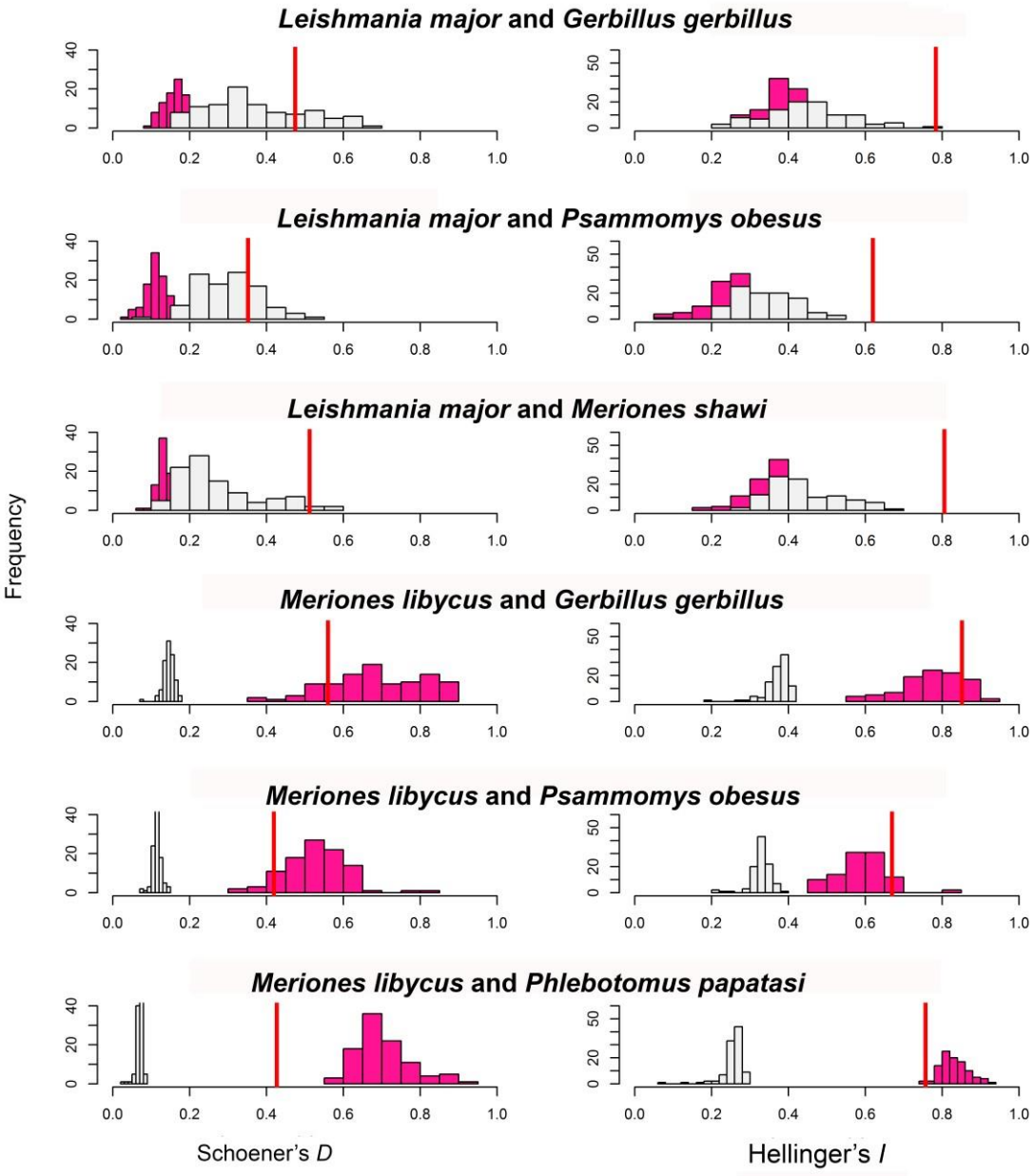
S3 File: Values of niche breadth for *Leishmania major*, *Phlebotomus papatasi*, and the four potential mammal reservoirs.

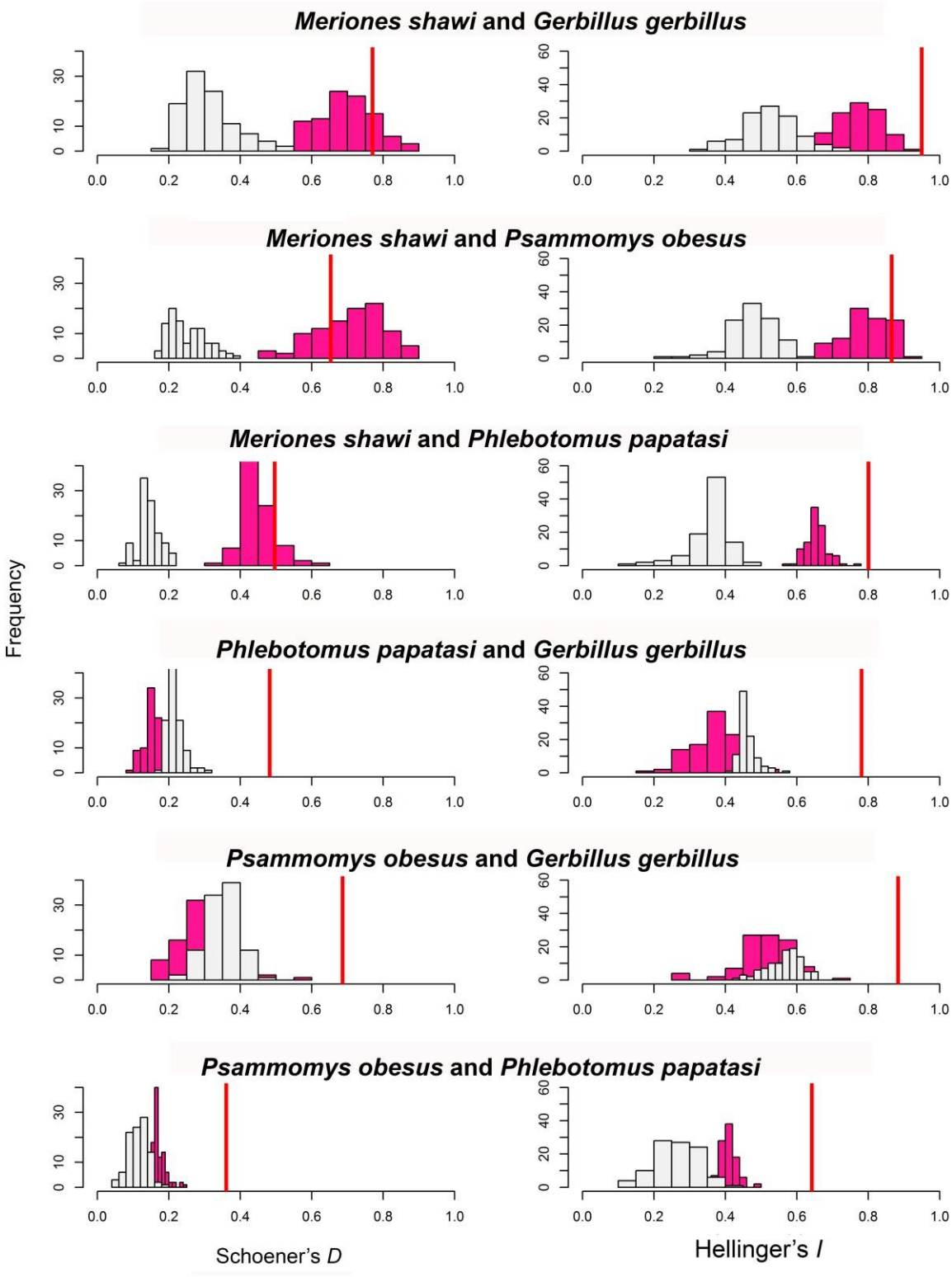
Species	Niche breadth
Leishmania major	0.097
Phlebotomus papatasi	0.165
Meriones libycus	0.283
Gerbillus gerbillus	0.092
Meriones shawi	0.246
Psammomys obesus	0.317

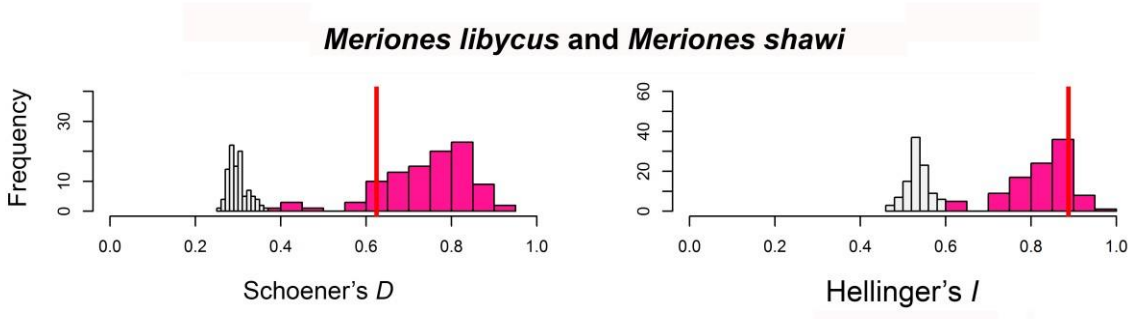
S4 File: Visualizations of ecological niches of four potential mammal reservoirs in two environmental dimensions. The diagram shows the overall environment available across Libya (gray), and the suitable conditions for species occurrences (pink).



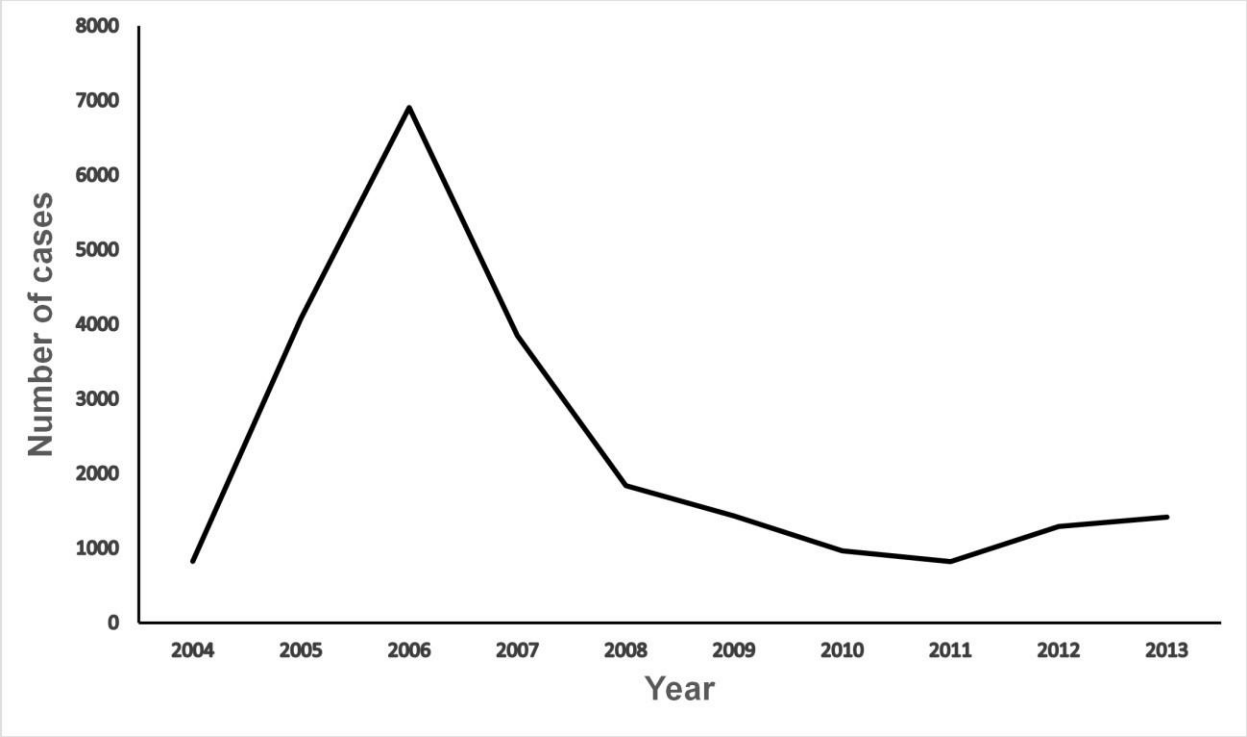
S5 File: Background similarity tests of ecological niche overlap between species. The red vertical line represent the observed niche overlap between the two ENMs in the question. The results of the background similarity tests were based on Schoener's *D* (left column) and Hellinger's *I* (right column) similarity metrics.



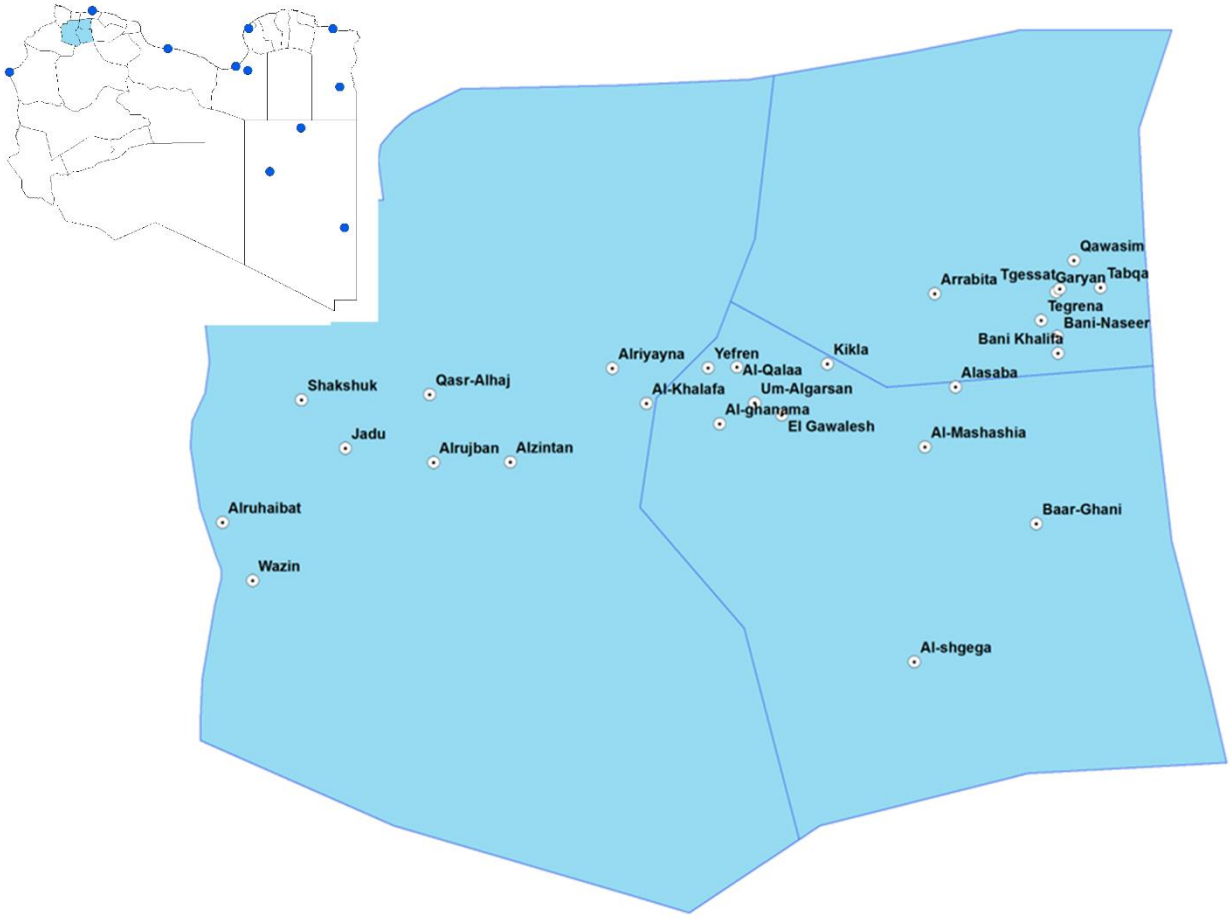

Meriones shawi and Gerbillus gerbillus
Meriones shawi and Psammomys obesus
Meriones shawi and Phlebotomus papatasi
Phlebotomus papatasi and Gerbillus gerbillus
Psammomys obesus and Gerbillus gerbillus
Psammomys obesus and Phlebotomus papatasi
Frequency
Schoener's D
Hellinger's I



S6 File: Total annual number of cases reported to the Libyan National Centre for Disease Control 2004-2013. These cases were reported by the local health units in each province and notified to the center for control measures based on the endemic status of each focus. These cases were identified by passive surveillance, and were not diagnosed to the species level.



S7 File: Localities with high zoonotic cutaneous leishmaniasis incidence and water resource management across Libya. Districts with high incidence are shown in blue, and localities within each district is presented as a dotted points. The map of Libya at the top shows the distribution of areas with water resource management initiatives as blue circles.



Chapter 3: Climate change influences on the global potential distribution of bluetongue virus[3]

[3] Samy AM, Peterson AT (2016). Climate change influences on the global potential distribution of bluetongue virus. PLoS One. 11(3):e0150489. doi: 10.1371/journal.pone.0150489

Abstract

The geographic distribution of arboviruses has received considerable attention after several dramatic emergence events around the world. Bluetongue virus (BTV) is classified among category "A" diseases notifiable to the World Organization of Animal Health (OIE), and is transmitted among ruminants by biting midges of the genus *Culicoides*. Here, we developed a comprehensive occurrence data set to map the current distribution, estimate the ecological niche, and explore the future potential distribution of BTV globally using ecological niche modeling and based on diverse future climate scenarios from general circulation models (GCMs) for four representative concentration pathways (RCPs). The broad ecological niche and potential geographic distribution of BTV under present-day conditions reflected the disease's current distribution across the world in tropical, subtropical, and temperate regions. All model predictions were significantly better than random expectations. As a further evaluation of model robustness, we compared our model predictions to 331 independent records from most recent outbreaks from the Food and Agriculture Organization Emergency Prevention System for Transboundary Animal and Plant Pests and Diseases Information System (EMPRES-i); all were successfully anticipated by the BTV model. Finally, we tested ecological niche similarity among possible vectors and BTV, and could not reject hypotheses of niche similarity. Under future-climate conditions, the potential distribution of BTV was predicted to broaden, especially in central Africa, United States, and western Russia.

Introduction

The global distribution of arboviruses has received considerable attention from public health organizations after recent emergence events in several parts of the world [1,2]. Bluetongue virus (BTV) is an arboviral disease in ruminants [3], caused by a member of the genus *Orivirus* in the family *Reoviridae*. The disease is transmitted among ruminants by the bites of biting midges of the genus *Culicoides* [4].

BTV has been responsible for massive sheep mortality; for example, outbreaks in the Mediterranean region since 1998 resulted in deaths of over 800,000 sheep [5]. A single strain of BTV in Belgium disrupted animal trade and killed animals with a market value of UK £180 million during a 2006-2007 outbreak [6]. In the United States, BTV causes losses of US $125 million yearly [7]. Previous reports have discussed early introduction of infected sheep into South Africa 125 years ago [8], but others identified South Africa as the origin of the infection [9].

BTV geography was long limited to a range between 40°N and 35°S [1]. Recently, however, several BTV strains began to spread worldwide [1,10,11], including to more northern parts of Europe, in 1998 [4]. The expansion and the potential for susceptibility of new vector species to the virus raises concerns of broader BTV spread [4,12-14]. BTV is transmitted by several vector species: *Culicoides imicola* Kieffer, 1913 (Diptera: Ceratopogonidae) is the most significant vector in the Old World [15], but three other species serve as vectors in the United Sates alone [16]. The distribution and movement of hosts has also been identified as a limiting factor for BTV spread; although BTV is known to have infected several ruminants, cattle and sheep are identified as primary reservoirs in several endemic areas worldwide [1]. The

combination of climate, presence of susceptible host, and presence of competent vectors marks areas where BTV can circulate in the long term among livestock.

Previous studies have mapped BTV risk based on occurrence data from single countries [17-19]; others included vector distributions in mapping efforts [5,18]. One study explored the global distribution and possible future shifts in the distribution of *C. imicola* across the world [15]. Ecological niche models provide a robust approach by which to assess and evaluate distribution of disease risk [20]: this approach has been used in mapping everything from fungal to arboviral diseases in several recent analyses [21,22].

Here, we developed a comprehensive database of BTV case occurrences, and estimated the global potential distribution of BTV under both current and future climate conditions. The study used outputs from 62 general climate models (GCMs) and four representative concentration pathway (RCP) scenarios from the Fifth Assessment Report (AR5) of the Intergovernmental Panel on Climate Change (IPCC) to estimate the future potential distribution of the virus. Finally, we tested niche similarity between several vector species and BTV case distribution in different geographic areas to provide some level of assessment of the role of particular potential vector species in BTV transmission.

Materials and Methods

Input data

Primary records of BTV occurrences (i.e. data reports of animal infections) were obtained from the PubMed database and Web of Knowledge using the search term "bluetongue virus," as well as from OIE reports (www.oie.int), the ReoID database (www.reoviridae.org/), and the Food and Agriculture Organization Emergency Prevention System for Transboundary Animal

and Plant Pests and Diseases Information System (EMPRES-i; http://empres-i.fao.org). Data regarding BTV occurrences through November 2014 were used in calibration, whereas occurrences from after that date (through September 2015) were used to provide a semi-independent data set for model evaluation (see Discussion). BTV records were drawn from diverse sources as we are seeking a global map of disease across the world; however, OIE and FAO data are limited to countries where BTV is notifiable, with most sampling in Europe and United States; data did not include the older BTV outbreaks in Africa and Asia. For niche comparisons with possible vectors, we collected vector occurrences from the Global Biodiversity Information Facility (GBIF; www.gbif.org) and literature in the PubMed and Web of Knowledge databases. The vector occurrences included records for six species: *C. imicola*, *C. insignis* Lutz, 1913, *C. variipennis* Coquillett, 1901, *C. sonorensis* Wirth & Jones, 1957, *C. occidentalis* Wirth & Jones, 1957, and *C. brevitarsis* Kieffer, 1917. When geographic references were textual, we assigned geographic coordinates based on consultation of online gazetteer data (www.gpsvisualizer.com). Data were filtered to eliminate duplicate records; we further reduced the data such that no pair of points was separated by <20 km (i.e., a single pixel) to reduce biases in calibrating ENMs [23]. The final occurrence data set was divided in two equal portions: half to calibrate the model, and half for evaluating model predictions.

To characterize current global climates, we used data available from the WorldClim archive (www.worldclim.org), which comprise19 bioclimatic variables derived from monthly temperature and rainfall values collected during 1950-2000 [24]. We used the 10′ spatial resolution in light of the global extent of our modeling efforts. For future conditions, we obtained data based on GCM outputs for 2050. These data comprised four RCPs spanning broadly different emissions scenarios into the future. Our future-climate projections thus summarized 62

combinations (S1 File). We used bioclimatic variables derived from monthly temperature and precipitation values because they are known factors in BTV transmission risk [1,25,26].

We omitted bioclimatic variables 8-9 and 18-19 from analysis, in light of known spatial artifacts in those four variables. The remaining of 15 variables were subjected to a principal components analysis (PCAs) to reduce the dimensionality of our models and avoid multicollinearity of variables (see summary of variable correlations in S2 File). The component loadings in the present-day data were used to transform future-climate data, using the ENMGadgets package [27] in R version 3.2.0 [28].

Ecological niche modeling

The maximum entropy algorithm implemented in Maxent version 3.3 [29] was used to estimate the ecological niche of BTV, roughly defined as the set of environmental conditions under which the species can maintain populations [20]. Our model was based on the first 6 principal components described above. We estimated the accessible area (**M**) [30,31] considering the geographic distribution of recent BTV outbreaks, which have been very broad, covering much of the world. We used the bootstrap functionality in Maxent to produce 100 replicate analyses. We used the median values across all models and replicates as a best estimate of the ecological niche of BTV. Finally, we calculated the median of the medians across all GCMs within each RCP scenario. Final models were thresholded based on a minimum allowable omission error rate of 5% (*E*=5%; [32]), assuming that a minimum of 5% of occurrences data may have errors in geolocation that misrepresented environmental values. We used the range (maximum – minimum) as an index of uncertainty between diverse models within each RCP.

Model robustness

Model robustness was evaluated using partial ROC statistics [20,32], which avoid many of the problems with traditional ROC approach [33]. We used the partialROC function in the ENMGadgets package in R [27] and the 50% subset of available occurrence data described above. A further evaluation of our model was based on independent data from recent outbreaks reported to FAO EMPRES-i. These data represent outbreaks reported between December 2014 and September 2015; that is, the evaluation data come from the year following the temporal span of the data used for model calibration. We used a one-tailed cumulative binomial probability distribution that assessed the probability of obtaining the observed level of correct prediction by a chance alone, given the background expectation of correct predictions based on the proportional coverage of the region by the thresholded model prediction.

Niche overlap of bluetongue virus and its vectors

We tested the niche similarity between each potential vector species and BTV using the background similarity test implemented in ENMTools version 1.4.4 [34]. We developed a specific **M** hypothesis [31] for each vector species as follows: *C. imicola*, a vector of BTV in the Old World [1,5,35], so we estimated a broad accessible area (**M**) that included all of Europe, Asia, and Africa for that species. *Culicoides insignis* is reported from North, Central, and South America [36], so its **M** was estimated to include all of the Americas. The **M** hypotheses for *C. variipennis* and *C. sonorensis* were estimated as all of North and Central America. *Culicoides brevitarsis* was restricted to East Asia and Australia, and *C. occidentalis* was limited to the southern United States and Central America.

The background similarity test assessed whether vector and BTV niches are less similar than expected given the "background" similarity manifested across the accessible areas of each [34]. We compared niche model similarity values based on actual occurrences of each species, with distributions of background similarity based on comparison of the niche of one species with "niche" models based on random points from across the **M** of the other species. We used numbers of random points equal to the number of actual occurrences for the other species. The null hypothesis of niche similarity was rejected if the observed *D* or *I* values for the BTV and vector species in question fell below the 5^{th} percentile in the random-replicate distribution.

Results

We assembled a total of 1677 unique occurrences for BTV around the world for model calibration. These points were filtered down to 1260 records in individual pixels. The overall pattern of occurrences indicated a geographically broad distribution of BTV, with more intense sampling efforts in Europe, where the virus invaded recently (Fig 1). Sampling was much more sparse in Africa and South America. Most BTV records were outside the early geographic belt identified for BTV distribution [37] (Fig 1). We also assembled an overall total of 798 occurrence records for six vector species: *C. imicola* (*N* = 408), *C. sonorensis* (*N* = 239), *C. variipennis* (*N* = 75), *C. insignis* (*N* = 33), *C. brevitarsis* (*N* = 23), and *C. occidentalis* (*N* = 20). These species have different ranges across the world (S3 File): *C. imicola* has a broader distribution extending from East Asia to western Africa; however, other species are limited in their ranges to East Asia and Australia (*C. brevitarsis*), North and Central America (*C.*

variipennis and *C. sonorensis*), southern United States and Central America (*C. occidentalis*), and North and South America (*C. insignis*).

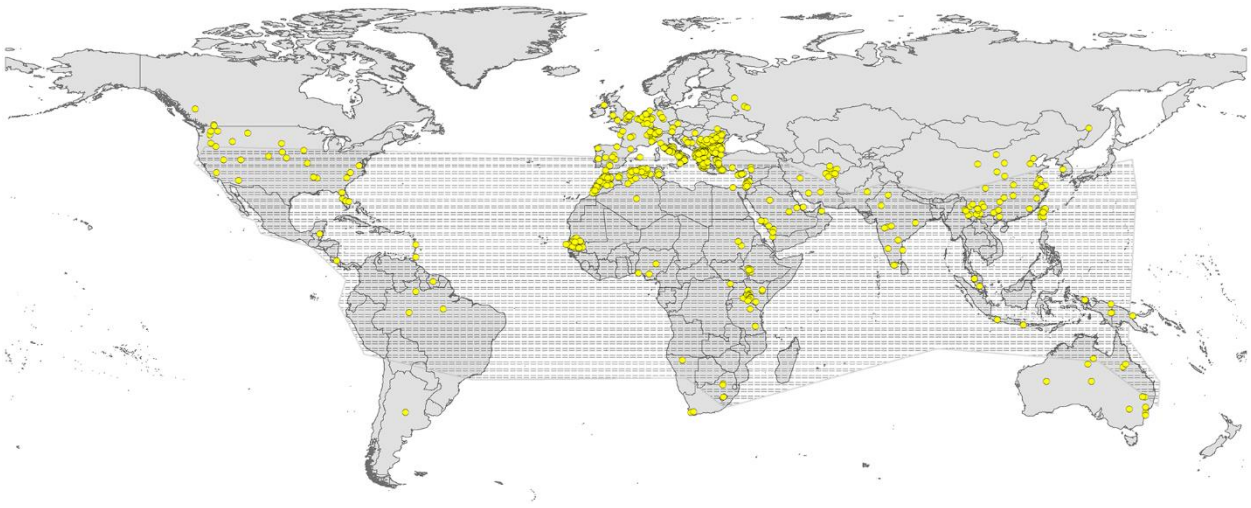


Fig 1: Summary of bluetongue virus occurrences (yellow points) available for model calibration worldwide. Dotted black shading represents the early belt of BTV occurrence.

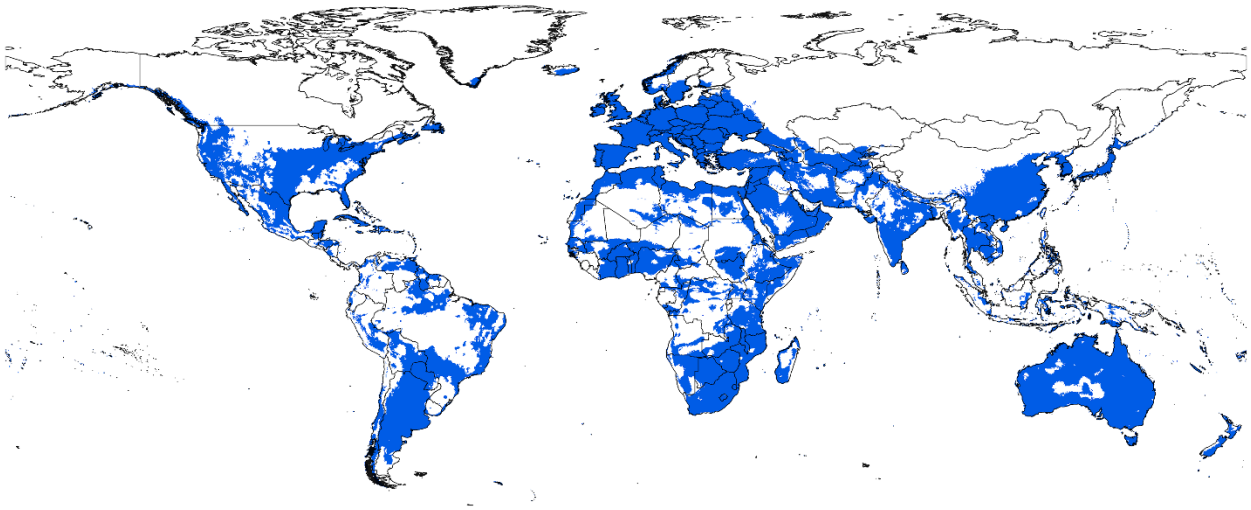


Fig 2: Current potential distribution map for bluetongue virus based on present-day climatic conditions. Blue shaded areas are modeled suitable conditions, and white areas are unsuitable conditions.

The potential distribution of BTV under present-day conditions showed high suitability across southern Europe, Australia, the Indian Subcontinent, and northern and southern Africa (Fig 2). BTV occurred in tropical, subtropical, and temperate climate zones. Suitable areas were also identified in West Africa, United States, and southern and western Canada. In all, the model outputs corresponded well to known areas of transmission around the world. Model predictions were significantly better than random expectations, in that partial ROC AUC ratios were uniformly higher than the random classifier with an AUC ratio of 1 ($P < 0.01$). The data set of 331 independent records was used to evaluate the robustness of our models in anticipating the current outbreaks across southern Europe, North Africa, and the United States. The model was significantly able to anticipate all 331 points reported for the most recent outbreaks of BTV (Fig 3; cumulative binomial test, $P < 0.0001$).

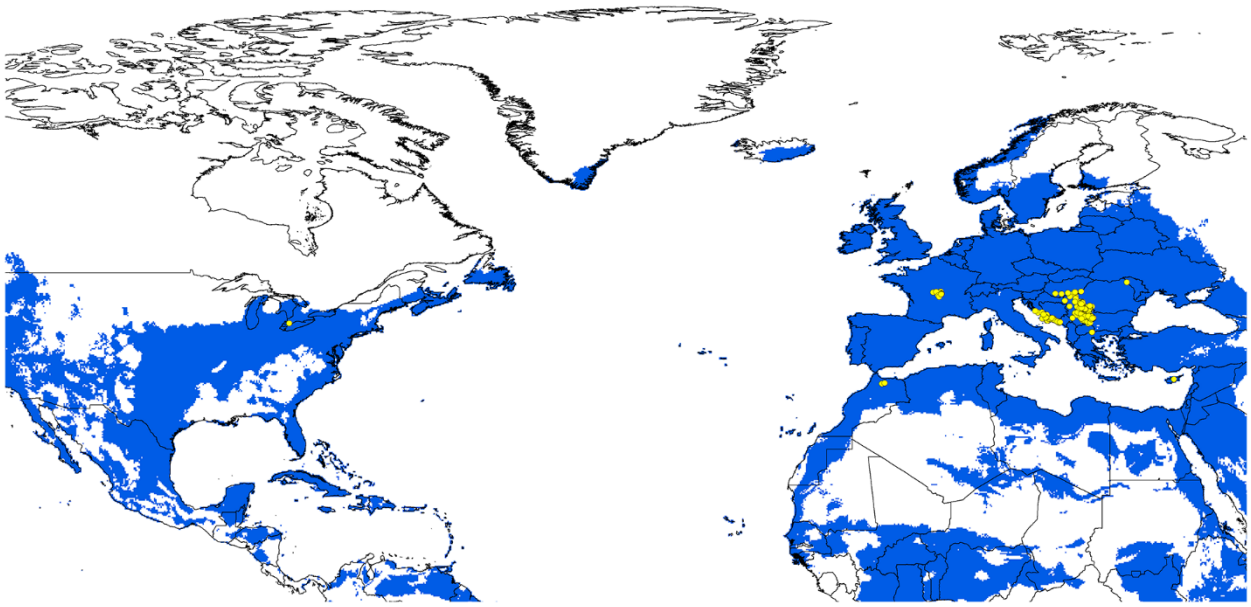


Fig 3: Relationship of additional independent BTV records to areas predicted as suitable for bluetongue virus occurrences. Yellow points are independent BTV occurrence data from the Old World and North America. Blue areas are represented as suitable and white as unsuitable.

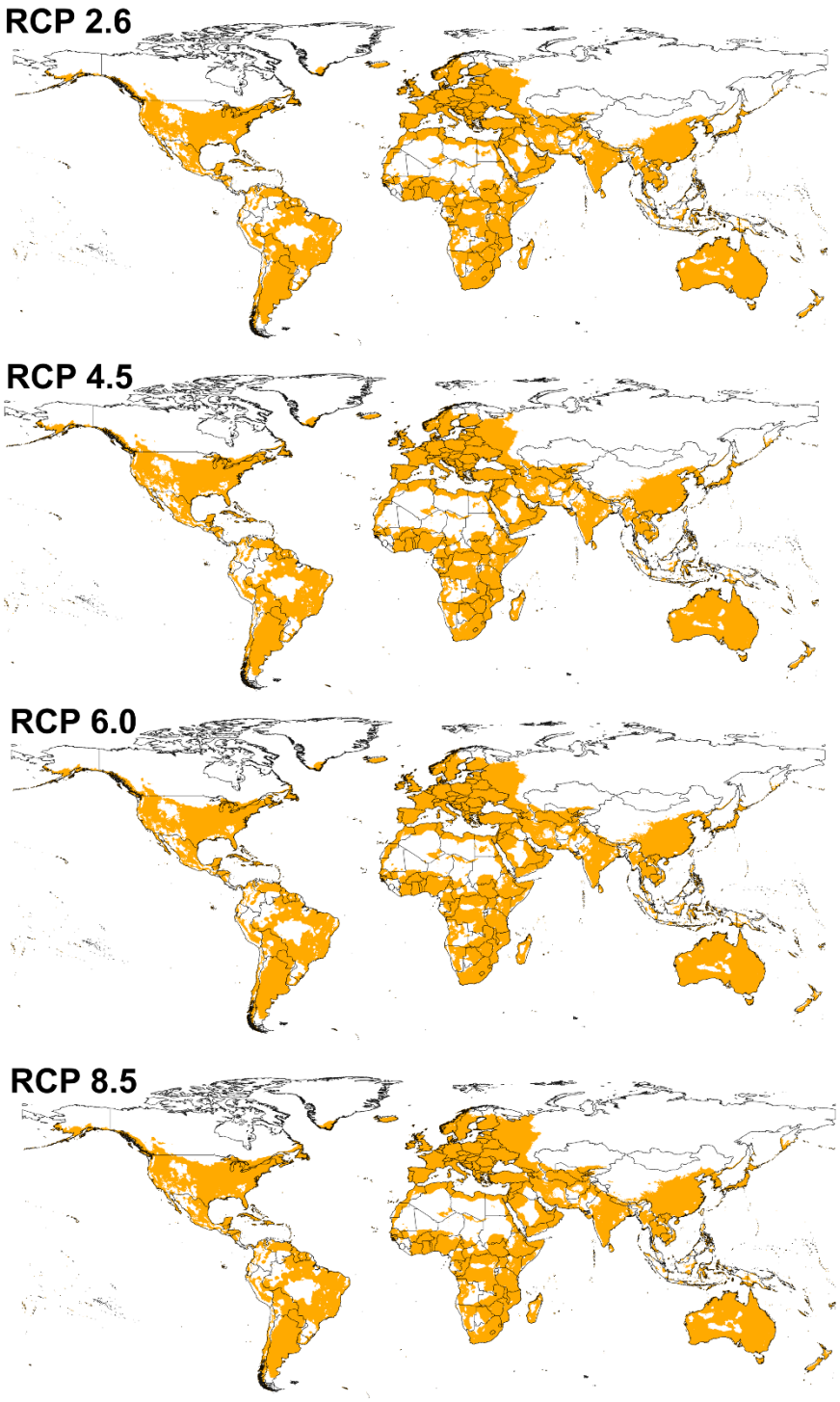


Fig 4: Predicted potential distribution maps for bluetongue virus under future climatic conditions. Models were calibrated across present-day conditions, and transferred to the future climate conditions. Each model is the median of all climate models across each representative concentration pathways (RCPs). Orange areas are modeled suitable conditions; white areas are unsuitable conditions for BTV occurrences.

BTV model transfers to future conditions indicated a pattern that was overall similar to that estimated for present-day conditions. However, the potential distribution of the virus under future conditions was broader, and included areas not identified as suitable under current conditions (Fig 4). This potential for expansion was particularly notable in central Africa, the United States, and western Russia. Under all future climate scenarios, the virus was seen to have a broader potential geographic distribution than at present (Fig 4). We noted few differences between GCMs within each climate scenario, such that model predictions were consistent over much of the world, with exceptions in western Russia, northern Europe, western South America, and Indonesia, where future projections were less stable (Fig 5).

BTV range increased from RCP 2.6 to RCP 8.5 (potential distributional area increased by 8.11% between present-day and RCP 2.6, and by 9.08% between present-day and RCP 8.5). Differences were also noted in the future potential BTV ranges of different models within each climate scenario (S4 File); the potential distributions under the different model conditions are summarized in the electronic supplementary materials as a GeoTIFF dataset (https://figshare.com/s/ac5383809b411c0f8779).

Finally, we tested the similarity of estimated niches between BTV and vector species, taking into account the background similarity between the accessible areas of each [34]. We could not reject the null hypothesis of niche similarity between BTV and any of the vector species ($P > 0.05$) (S5 File). Hence, present-day environmental conditions occupied by BTV and its vectors were not demonstrably different, and we found no indication of unlinked transmission and vector occurrence.

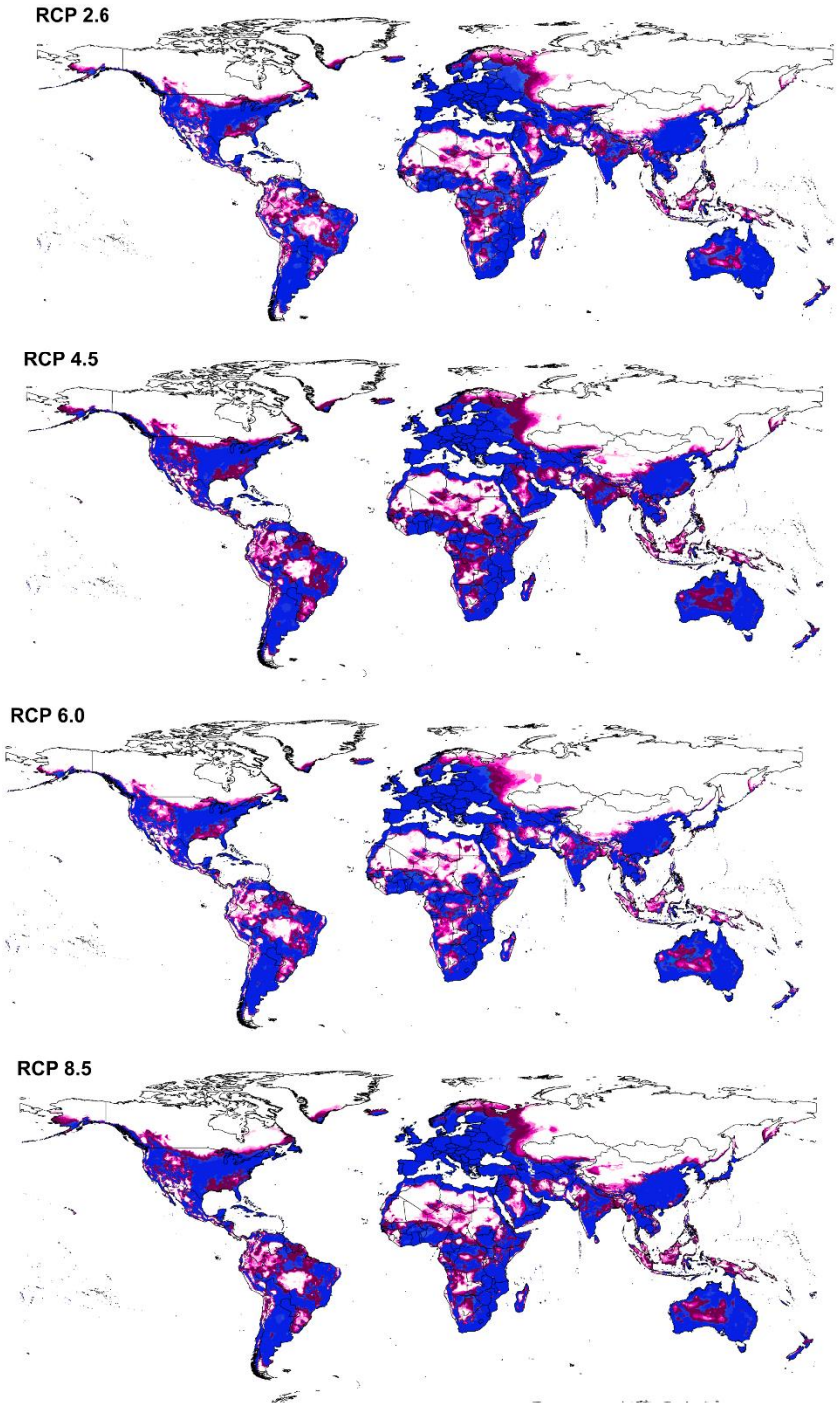


Fig 5: Summary of the modeled global distribution of bluetongue virus under both current and future climatic conditions to show the stability of predictions at present and into the future, and to illustrate differences among representative concentration pathways (RCPs). Dark blue represents model stability under both current and future conditions, light blue represents low agreement between current and future conditions, dark purple represents agreement among all climate models in anticipating potential distributional areas in the future, and light purple indicates low agreement between diverse climate models as regards distributional potential in the future.

Discussion

This study built a comprehensive database of BTV occurrences from 1964 through November 2014. We used these data to map the potential distribution of BTV in the present-day, and also to identify future potential distributional shifts in view of the most recent future climate scenario model outputs. The availability of BTV data was poor until large-scale outbreaks across Europe began in the late 1990s; since then, the disease has caused serious impacts on international animal trade, and serious illness and high mortality rates among ruminants [38,39], and reporting has been more detailed.

Climate has been suggested as a major driver of the distribution of BTV [1]; for example, the European outbreaks were thought to be a consequence of warming climates [1], and the virus expanded another ~800 north in 2005 [40,41]. Historically, the disease was found in a belt between 40ºN and 35ºS that included northern Australia, parts of the Indian Subcontinent, Middle East, Africa, Cyprus, and the Americas. Prior to that point, BTV was known only from South Africa and Cyprus [8, 42], The origin of the disease thus has three possible explanations: (1) BTV was present in both Africa and Europe but was not documented owing to misidentification or poor diagnostic tools; (2) BTV originated in South Africa and dispersed to Europe; (3) BTV originated in Europe and dispersed to South Africa. Phylogenetic analyses suggest that strains responsible for the new BTV outbreaks are similar to those that circulated for decades in early epidemic sites [1]. FAOSTAT data indicate results revealed continuous international trade between Europe and Africa [43], which could allow viral infections in animals to move between the two continents. This observation also suggests that the virus will be able to spread to new sites that become suitable as climatic conditions change in coming decades.

We modelled BTV occurrences using the most recent version of 62 future climate model outputs, and used four RCPs to summarize variations among possible greenhouse gas concentration trajectories. Our results indicated the possibility of range expansion to other regions where the virus is not presently endemic, in response to climate changes [1] .

Previous studies predicted the distribution of BTV and its vectors; most of these studies were limited to Europe after BTV emergence in Europe [5,18,26], and assessed transmission risk based mostly on vector distribution and abundance [5,18]. However, a single study took the advantage of a mechanistic models to quantify R_0 values of BTV across Europe based on different temporal scales of climate data. This latter study demonstrated BTV risk areas across most of Europe that coincide with our results. However, our prediction covers a much broader portion of Europe, extending across much of the continent, east to western Russia. A recent study of BTV potential in the United States predicted that further northwards expansion of *C. sonorensis* can be expected in the future [44]. Our current study anticipated both the northern United States, southwestern Canada, and Ontario as at risk, as long as both BTV and vector expand distributionally in tandem; however, it marked unsuitability of conditions to virus spread along the US Gulf Coast and in the eastern Rocky Mountains, where livestock and wild animals have been diagnosed as positive to BTV [10]. This discord may reflect limitations of our model; however, these same regions were identified as suitable in our future models, which may indicate that our present-day models cannot anticipate risk in these regions given ongoing climate reorganization.

The models estimated in our study showed significant performance in two tests based on testing data that are ostensibly independent from the data used in model calibration. However, we see some room for concern about this level of independence because disease events are

dependent events on a number of levels—individual cases may be linked to one another via pathogen and vector population biology, on both micro and macro scales. Similarly, surveillance is often responsive, and gets concentrated in affected areas, which can create additional dependencies. For these reasons, our data set was selected to constitute different data sources that represent sampling across the world and not just notified data from EMPRES-i. Our evaluation tested the possibility of these models to predict the recent outbreaks across the world from December 2014 until the most recent outbreak in Ontario, Canada. These models could anticipate all of these current outbreaks.

Different levels of uncertainty were associated with the mapping process (S6 File). These uncertainty estimates were based on variations in predicted distributions of BTV across different climate models rather than just an estimate of internal uncertainty for predictions under the same climate model [22,45].

Finally, our models offered an interesting perspective on vector associations of BTV infections around the world. The BTV transmission cycle includes hosting by ruminants and vectors that transmit virus between hosts. Major knowledge gaps include the broader host distribution of BTV in diverse livestock hosts, and the vanishingly few studies that have focused on vector competence of different species of *Culicoides*. This study characterized diversity in the species of vectors associated with BTV in different parts of the world: *C. imicola* in the Old World, *C. sonorensis* and *C. variipennis* in North America, *C. insignis* in North and South America, *C. occidentalis* in southern United States and Central America, and *C. brevitarsis* in Australia. We tested niche similarity between each vector and BTV distribution across the accessible region for the corresponding vector; we could not reject the null hypothesis in any case, so vector populations and BTV appear to share similar ecological niches. This observation

is important because (1) vector populations can assist in identifying the potential distribution of the BTV in countries where the disease is not reported dependably to authorities, and (2) vector populations may drive BTV response to climate change [46].

The current study leaves important questions unanswered regarding the global distribution of BTV: one related to the relationship between vector population dynamics and changes in BTV transmission, host response to climate change, and responses of different BTV serotypes to climate and how much these responses are similar or different. Our future work will focus on exploring these questions to illuminate additional key details of BTV epidemiology and ecology around the world.

Acknowledgments

AMS was funded by the Graduate Egyptian Fulbright Mission Program (EFMP). The authors thank the KU ENM group, and the Department of Entomology and the Research and Training Center on Vectors of Diseases of Ain Shams University, for their support during this work. Thanks also go to Qiao Huijie for his support and assistance during the study.

References

1. Purse BV, Mellor PS, Rogers DJ, Samuel AR, Mertens PP, Baylis M (2005) Climate change and the recent emergence of bluetongue in Europe. Nat Rev Microbiol 3: 171-181.
2. Dash AP, Bhatia R, Sunyoto T, Mourya DT (2013) Emerging and re-emerging arboviral diseases in Southeast Asia. J Vector Borne Dis 50: 77-84.

3. Maclachlan NJ, Mayo CE (2013) Potential strategies for control of bluetongue, a globally emerging, *Culicoides*-transmitted viral disease of ruminant livestock and wildlife. Antiviral Res 99: 79-90.

4. Wilson AJ, Mellor PS (2009) Bluetongue in Europe: past, present and future. Philos Trans R Soc Lond B Biol Sci 364: 2669-2681.

5. Tatem AJ, Baylis M, Mellor PS, Purse BV, Capela R, Pena I, et al. (2003) Prediction of bluetongue vector distribution in Europe and North Africa using satellite imagery. Vet Microbiol 97: 13-29.

6. Webb D (2008) The economic and social impact of the Institute for Animal Health's work on bluetongue disease (BTV-8). Pirbright Institute. Available at http://WWW/ecosoc/docs/Blue-Tongue-case-studypdf.

7. Tabachnick WJ (1996) *Culicoides variipennis* and bluetongue-virus epidemiology in the United States. Annu Rev Entomol 41: 23-43.

8. Walton TE (2004) The history of bluetongue and a current global overview. Vet Ital 40: 31-38.

9. Gerdes GH (2004) A South African overview of the virus, vectors, surveillance and unique features of bluetongue. Vet Ital 40: 39-42.

10. Ruder MG, Lysyk TJ, Stallknecht DE, Foil LD, Johnson DJ, Chase CC, et al. (2015) Transmission and epidemiology of bluetongue and epizootic hemorrhagic disease in North America: Current perspectives, research gaps, and future directions. Vector Borne Zoonotic Dis 15: 348-363.

11. McVey DS, MacLachlan NJ (2015) Vaccines for prevention of bluetongue and epizootic hemorrhagic disease in livestock: A North American perspective. Vector Borne Zoonotic Dis 15: 385-396.

12. Sreenivasulu D, Subba Rao MV, Reddy YN, Gard GP (2004) Overview of bluetongue disease, viruses, vectors, surveillance and unique features: The Indian Sub-continent and adjacent regions. Vet Ital 40: 73-77.

13. Lager IA, Duffy S, Miquet J, Vagnozzi A, Gorchs C, Draghi M, et al. (2004) Incidence and isolation of bluetongue virus infection in cattle of the Santo Tome Department, Corrientes Province, Argentina. Vet Ital 40: 141-144.

14. Legisa DM, Gonzalez FN, Dus Santos MJ (2014) Bluetongue virus in South America, Central America and the Caribbean. Virus Res 182: 87-94.

15. Guichard S, Guis H, Tran A, Garros C, Balenghien T, Kriticos DJ (2014) Worldwide niche and future potential distribution of *Culicoides imicola*, a major vector of bluetongue and African horse sickness viruses. PLoS ONE 9: e112491.

16. Pfannenstiel RS, Mullens BA, Ruder MG, Zurek L, Cohnstaedt LW, Nayduch D (2015) Management of North American *Culicoides* biting midges: Current knowledge and research needs. Vector Borne Zoonotic Dis 15: 374-384.

17. Torina A, Caracappa S, Mellor PS, Baylis M, Purse BV (2004) Spatial distribution of bluetongue virus and its *Culicoides* vectors in Sicily. Med Vet Entomol 18: 81-89.

18. Racloz V, Presi P, Vounatsou P, Schwermer H, Casati S, Vanzetti T, et al. (2007) Use of mapping and statistical modelling for the prediction of bluetongue occurrence in Switzerland based on vector biology. Vet Ital 43: 513-518.

19. Racloz V, Venter G, Griot C, Stark KD (2008) Estimating the temporal and spatial risk of bluetongue related to the incursion of infected vectors into Switzerland. BMC Vet Res 4: 42.

20. Peterson AT, Soberón J, Pearson RG, Anderson RP, Martínez-Meyer E, Nakamura M, et al. (2011) Ecological Niches and Geographic Distributions Princeton: Princeton University.

21. Samy AM, van de Sande WW, Fahal AH, Peterson AT (2014) Mapping the potential risk of mycetoma infection in Sudan and South Sudan using ecological niche modeling. PLOS Negl Trop Dis 8: e3250.

22. Pigott DM, Bhatt S, Golding N, Duda KA, Battle KE, Brady OJ, et al. (2014) Global distribution maps of the leishmaniases. Elife. 3: e02851.

23. Syfert MM, Smith MJ, Coomes DA (2013) The effects of sampling bias and model complexity on the predictive performance of MaxEnt species distribution models. PLoS ONE 8: e55158.

24. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. Int J Climatol 25: 1965-1978.

25. Faes C, van der Stede Y, Guis H, Staubach C, Ducheyne E, Hendrickx G, et al. (2013) Factors affecting bluetongue serotype 8 spread in northern Europe in 2006: The geographical epidemiology. Prev Vet Med 110: 149-158.

26. Guis H, Caminade C, Calvete C, Morse AP, Tran A, Baylis M (2012) Modelling the effects of past and future climate on the risk of bluetongue emergence in Europe. J R Soc Interface 9: 339-350.

27. Barve N, Barve V (2013) ENMGadgets: Tools for Pre and Post Processing in ENM Workflows. Available at https://github.com/vijaybarve/ENMGadgets.

28. R Development Core Team (2015) A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austia. Available at http://www.R-project.org.

29. Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecol Model 190: 231-259.

30. Soberón J, Peterson AT (2005) Interpretation of models of fundamental ecological niches and species' distributional areas. Biodivers Inform: 1-10.

31. Barve N, Barve V, Jiménez-Valverde A, Lira-Noriega A, Maher SP, Peterson AT, et al. (2011) The crucial role of the accessible area in ecological niche modeling and species distribution modeling. Ecol Model 222: 1810-1819.

32. Peterson AT, Papeş M, Soberón J (2008) Rethinking receiver operating characteristic analysis applications in ecological niche modeling. Ecol Model 213: 63-72.

33. Lobo JM, Jiménez-Valverde A, Real R (2008) AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography 17: 145-151.34.

34. Warren DL, Glor RE, Turelli M (2008) Environmental niche equivalency versus conservatism: Quantitative approaches to niche evolution. Evol 62: 2868-2883.

35. Patakakis MJ, Papazahariadou M, Wilson A, Mellor PS, Frydas S, et al. (2009) Distribution of *Culicoides* in Greece. J Vector Ecol 34: 243-251.

36. Tanya VN, Greiner EC, Gibbs EP (1992) Evaluation of *Culicoides insignis* (Diptera: Ceratopogonidae) as a vector of bluetongue virus. Vet Microbiol 32: 1-14.

37. Tabachnick WJ (2010) Challenges in predicting climate and environmental effects on vector-borne disease episystems in a changing world. J Exp Biol 213: 946-954.

38. MacLachlan NJ (2004) Bluetongue: Pathogenesis and duration of viraemia. Vet Ital 40: 462-467.

39. Turner J, Bowers RG, Baylis M (2012) Modelling bluetongue virus transmission between farms using animal and vector movements. Sci Rep 2: 319.

40. Radostits OM, Gay CC, Hinchcliff KW, Constable PD (2007) Veterinary Medicine: A Textbook of the Diseases of Cattle, Sheep, Pigs, Goats and Horses Tenth Edition. Saunders Ltd, Edinburgh, UK.

41. Venter GJ, Mellor PS, Paweska JT (2006) Oral susceptibility of South African stock-associated *Culicoides* species to bluetongue virus. Med and Vet Entomol 20: 329-334.

42. Verwoerd DW (2009) History of bluetongue research at Onderstepoort. Onderstepoort J Vet Res 76: 99-102.

43. Food and Agriculture Organization of the United Nations (2015) FAOSTAT database. Available at http://faostat3.fao.org/home/E.

44. Zuliani A, Massolo A, Lysyk T, Johnson G, Marshall S, Berger K, et al. (2015) Modelling the northward expansion of *Culicoides sonorensis* (Diptera: Ceratopogonidae) under future climate scenarios. PLoS ONE 10: e0130294.

45. Kraemer MU, Sinka ME, Duda KA, Mylne AQ, Shearer FM, Barker CM, et al. (2015) The global distribution of the arbovirus vectors *Aedes aegypti* and *Ae. albopictus*. Elife. 4: e08347.

46. Mouchet J, Carnevale P (1997) Impact of changes in the environment on vector-transmitted diseases. Sante 7: 263-269.

Supporting Information

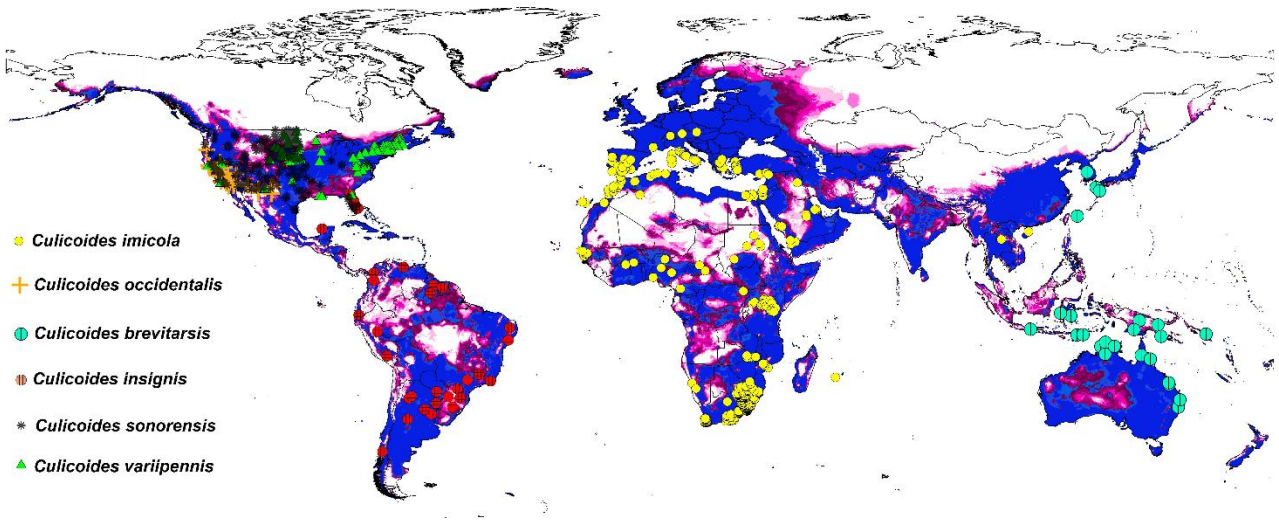
S1 File: A summary of four representative concentration pathways and 62 climate models used in BTV model projection in future climate conditions.

RCP 2.6	RCP 4.5	RCP 6.0	RCP 8.5
BCC-CSM1-1	ACCESS1-0	BCC-CSM1-1	ACCESS1-0
CCSM4	BCC-CSM1-1	CCSM4	BCC-CSM1-1
CNRM-CM5	CCSM4	GFDL-ESM2G	CCSM4
GFDL-ESM2G	CESM1-CAM5-1-FV2	GISS-E2-R	CNRM-CM5
GFDL-CM3	CNRM-CM5	HadGEM2-AO	GFDL-CM3
GISS-E2-R	GFDL-ESM2G	HadGEM2-ES	GISS-E2-R
HadGEM2-AO	GFDL-CM3	IPSL-CM5A-LR	HadGEM2-AO
HadGEM2-ES	GISS-E2-R	MIROC5	HadGEM2-ES
IPSL-CM5A-LR	HadGEM2-AO	MRI-CGCM3	HadGEM2-CC
MIROC5	HadGEM2-ES	MIROC-ESM-CHEM	INMCM4
MRI-CGCM3	HadGEM2-CC	MIROC-ESM	IPSL-CM5A-LR
MIROC-ESM-CHEM	INMCM4	NorESM1-M	MIROC5
MPI-ESM-LR	IPSL-CM5A-LR		MRI-CGCM3
MIROC-ESM	MIROC5		MIROC-ESM-CHEM
NorESM1-M	MRI-CGCM3		MPI-ESM-LR
	MIROC-ESM-CHEM		MIROC-ESM
	MPI-ESM-LR		NorESM1-M
	MIROC-ESM		
	NorESM1-M		

S2 File: Correlation matrix showing patterns of relationships among environmental variables used in model calibration.

Variable	bio_1	bio_2	bio_3	bio_4	bio_5	bio_6	bio_7	bio_10	bio_11	bio_12	bio_13	bio_14	bio_15	bio_16	bio_17
bio_1		0.503988	0.840299	-0.83457	0.895538	0.967518	-0.73093	0.936067	0.979999	0.391349	0.464626	0.07566	0.359669	0.45428	0.102541
bio_2			0.371656	-0.19028	0.70028	0.33456	0.038541	0.609913	0.407114	-0.24751	-0.10919	-0.37226	0.510713	-0.12873	-0.36951
bio_3				-0.8896	0.61746	0.887752	-0.82298	0.665556	0.890835	0.571241	0.584606	0.239513	0.272658	0.585058	0.272086
bio_4					-0.51477	-0.93683	0.970513	-0.58932	-0.92637	-0.56021	-0.5708	-0.26078	-0.17903	-0.57233	-0.28971
bio_5						0.763464	-0.35802	0.987883	0.795494	0.133162	0.243795	-0.12216	0.426956	0.22644	-0.10359
bio_6							-0.87638	0.826259	0.995927	0.497649	0.534856	0.186373	0.260005	0.53004	0.215709
bio_7								-0.45784	-0.8466	-0.62015	-0.59144	-0.36054	-0.0575	-0.59742	-0.38911
bio_10									0.849946	0.215941	0.316141	-0.05537	0.40822	0.299732	-0.03435
bio_11										0.464706	0.518795	0.142479	0.308488	0.512126	0.171125
bio_12											0.89739	0.717207	-0.17678	0.922582	0.750824
bio_13												0.404306	0.132361	0.992763	0.440091
bio_14													-0.51566	0.439938	0.99403
bio_15														0.088541	-0.51312
bio_16															0.475818
bio_17															

S3 File: Summary of BTV vector occurrences available for testing niche similarity between BTV and vector niches and based on accessible area (M).



S4 File: Range of BTV expansion based on presence-absence matrix of each ecological niche model for corresponding climate model.

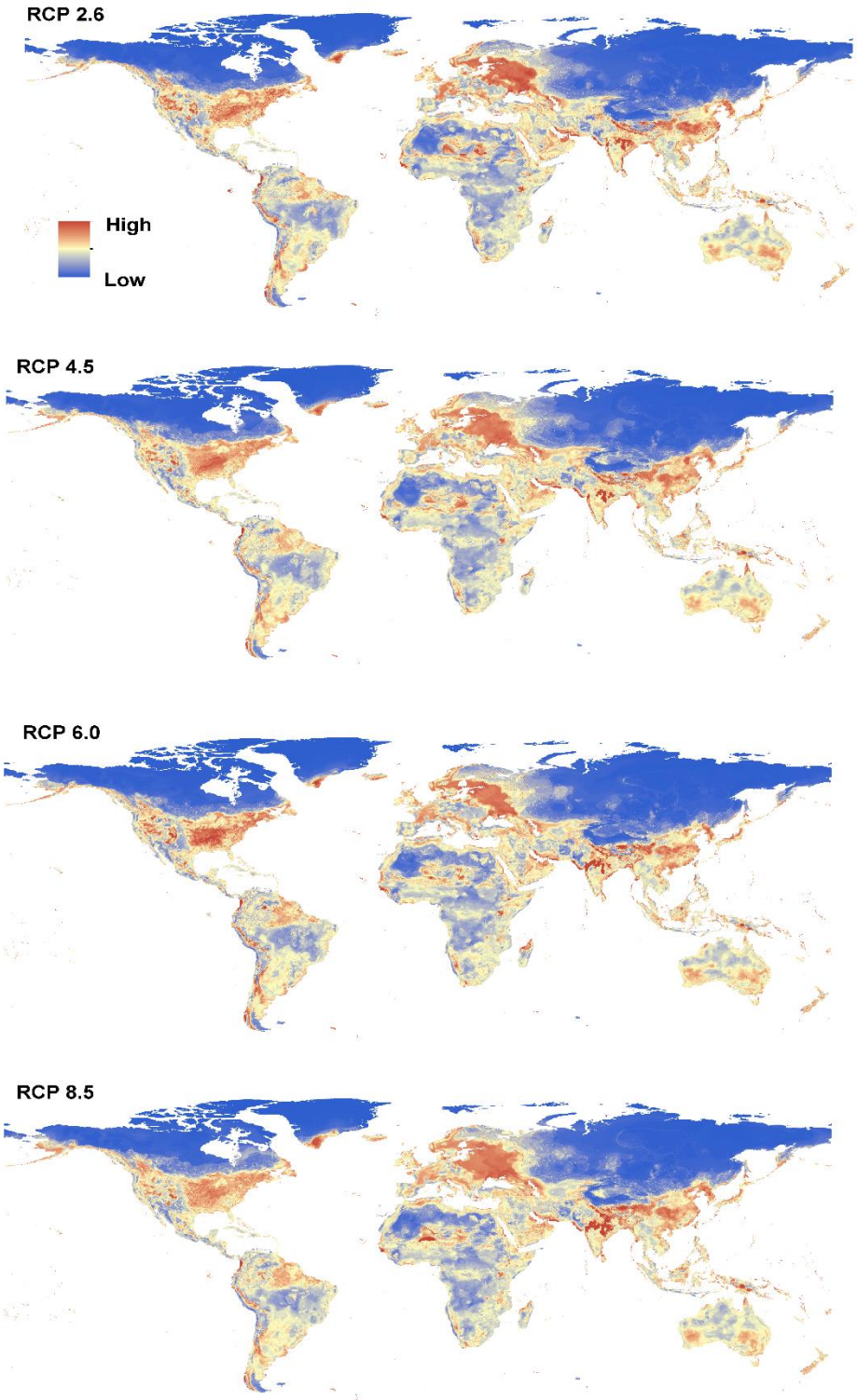
Scenario	Model	Absent	Present	Present/absent	Percentage
RCP2.6	BCC-CSM1-1	320366	264155	0.451917	45.1917
RCP2.6	CCSM4	317897	266624	0.456141	45.6141
RCP2.6	CNRM-CM5	325405	259116	0.443296	44.32963
RCP2.6	GFDL-ESM2G	334625	249896	0.427523	42.75227
RCP2.6	GFDL-CM3	304287	280234	0.479425	47.9425
RCP2.6	GISS-E2-R	323892	260629	0.445885	44.58848
RCP2.6	HadGEM2-AO	317299	267222	0.457164	45.71641
RCP2.6	HadGEM2-ES	320712	263809	0.451325	45.13251
RCP2.6	IPSL-CM5A-LR	318589	265932	0.454957	45.49571
RCP2.6	MIROC5	325238	259283	0.443582	44.3582
RCP2.6	MRI-CGCM3	329033	255488	0.43709	43.70895
RCP2.6	MIROC-ESM-CHEM	306733	277788	0.47524	47.52404
RCP2.6	MIROC-ESM	318977	265544	0.454293	45.42933
RCP2.6	NorESM1-M	332937	251584	0.430411	43.04105
RCP4.5	ACCESS1-0	314705	269816	0.461602	46.16019
RCP4.5	BCC-CSM1-1	335384	249137	0.426224	42.62242
RCP4.5	CCSM4	319042	265479	0.454182	45.41821
RCP4.5	CESM1-CAM5-1-FV2	285782	298739	0.511083	51.10834
RCP4.5	CNRM-CM5	313548	270973	0.463581	46.35813
RCP4.5	GFDL-ESM2G	318243	266278	0.455549	45.55491
RCP4.5	GFDL-CM3	298019	286502	0.490148	49.01483
RCP4.5	GISS-E2-R	324498	260023	0.444848	44.4848
RCP4.5	HadGEM2-AO	323918	260603	0.44584	44.58403
RCP4.5	HadGEM2-ES	312547	271974	0.465294	46.52938
RCP4.5	HadGEM2-CC	321261	263260	0.450386	45.03859
RCP4.5	INMCM4	318094	266427	0.455804	45.5804
RCP4.5	IPSL-CM5A-LR	300398	284123	0.486078	48.60783
RCP4.5	MIROC5	326305	258216	0.441757	44.17566
RCP4.5	MRI-CGCM3	338856	245665	0.420284	42.02843
RCP4.5	MIROC-ESM-CHEM	301046	283475	0.48497	48.49697
RCP4.5	MPI-ESM-LR	306650	277871	0.475382	47.53824
RCP4.5	MIROC-ESM	299136	285385	0.488237	48.82374
RCP4.5	NorESM1-M	330131	254390	0.435211	43.52111
RCP6.0	BCC-CSM1-1	326029	258492	0.442229	44.22288

RCP6.0	CCSM4	320166	264355	0.452259	45.22592
RCP6.0	GFDL-ESM2G	319455	265066	0.453476	45.34756
RCP6.0	GISS-E2-R	326015	258506	0.442253	44.22527
RCP6.0	HadGEM2-AO	309356	275165	0.470753	47.0753
RCP6.0	HadGEM2-ES	314539	269982	0.461886	46.18859
RCP6.0	IPSL-CM5A-LR	317246	267275	0.457255	45.72547
RCP6.0	MIROC5	325827	258694	0.442574	44.25743
RCP6.0	MRI-CGCM3	329559	254962	0.43619	43.61896
RCP6.0	MIROC-ESM-CHEM	297690	286831	0.490711	49.07112
RCP6.0	MIROC-ESM	315879	268642	0.459593	45.95934
RCP6.0	NorESM1-M	335711	248810	0.425665	42.56648
RCP8.5	ACCESS1-0	317911	266610	0.456117	45.61171
RCP8.5	BCC-CSM1-1	318566	265955	0.454996	45.49965
RCP8.5	CCSM4	316734	267787	0.458131	45.81307
RCP8.5	CNRM-CM5	325759	258762	0.442691	44.26907
RCP8.5	GFDL-CM3	282771	301750	0.516235	51.62347
RCP8.5	GISS-E2-R	315907	268614	0.459546	45.95455
RCP8.5	HadGEM2-AO	316666	267855	0.458247	45.8247
RCP8.5	HadGEM2-ES	315562	268959	0.460136	46.01357
RCP8.5	HadGEM2-CC	305062	279459	0.478099	47.80992
RCP8.5	INMCM4	318449	266072	0.455197	45.51966
RCP8.5	IPSL-CM5A-LR	296315	288206	0.493064	49.30636
RCP8.5	MIROC5	326152	258369	0.442018	44.20183
RCP8.5	MRI-CGCM3	318310	266211	0.455434	45.54344
RCP8.5	MIROC-ESM-CHEM	308618	275903	0.472016	47.20155
RCP8.5	MPI-ESM-LR	307478	277043	0.473966	47.39659
RCP8.5	MIROC-ESM	308948	275573	0.471451	47.1451
RCP8.5	NorESM1-M	324077	260444	0.445568	44.55683

S5 File: Results of background similarity tests assessing niche similarity between bluetongue virus and six vector species. The null hypothesis of niche similarity was rejected if the observed D or I values for the BTV and vector species in question fell below the 5th percentile in the random-replicate distribution (i.e. 5% in table).

	Schoener's *D*				Hellinger's *I*			
Species	Observed	5%	95%	*P* value	Observed	5%	95%	*P* value
Culicoides imicola	0.585	0.466	0.863	*P > 0.05*	0.847	0.677	0.926	*P > 0.05*
Culicoides insignis	0.615	0.382	0.683	*P > 0.05*	0.873	0.514	0.944	*P > 0.05*
Culicoides variipennis	0.618	0.419	0.804	*P* > 0.05	0.872	0.615	0.966	*P* > 0.05
Culicoides sonorensis	0.696	0.429	0.717	*P* > 0.05	0.923	0.598	0.958	*P* > 0.05
Culicoides brevitarsis	0.708	0.322	0.942	*P* > 0.05	0.927	0.551	0.956	*P* > 0.05
Culicoides occidentalis	0.505	0.233	0.739	*P* > 0.05	0.805	0.366	0.821	*P* > 0.05

S6 File: Uncertainty estimates associated with BTV mapping process in different climate models within each representative concentration pathway.



Chapter 4: Phylogeography and host transitions of Rift Valley Fever virus in Africa and the Arabian Peninsula

Abstract

Rift Valley Fever is an acute zoonotic viral disease caused by Rift Valley Fever virus (RVFV) that affects ruminants and humans in Sub-Saharan Africa and the Arabian Peninsula. We used phylogenetic analyses to understand the demographic history of RVFV populations, using sequence data from the three minigenomic segments of the virus. We used phylogeographic approaches to infer RVFV historical movement patterns across its geographic range, and to reconstruct transitions among host species. Results revealed broad circulation of the virus in East Africa, with many lineages originating in Kenya. Arrival of RVFV in Madagascar resulted from three major waves of virus introduction: the first from Zimbabwe, and the second and third from Kenya. The two major outbreaks in Egypt since 1977 possibly resulted from a long-distance introduction from Zimbabwe during the 1970s, and a single introduction took RVFV from Kenya to Saudi Arabia. Movement of the virus between Kenya and Sudan, and Central African Republic and Zimbabwe, was in both directions. Viral populations in West Africa appear to have resulted from a single introduction from Central African Republic. Finally, host transition analysis identified both humans and livestock as natural hosts of RVFV. The overall picture of RVFV history is thus one of considerable mobility, and dynamic evolution and biogeography, emphasizing its invasive potential, potentially more broadly than its current distributional limits.

Introduction

Rift Valley Fever is an acute zoonotic viral disease caused by Rift Valley Fever virus (RVFV; *Phlebovirus*, family *Bunyaviridae*) that affects both large mammals and humans, and that is transmitted by *Aedes* and *Culex* mosquitoes [1]. Rift Valley Fever causes high mortality and abortions in ruminants [2]; infections in humans are characterized by febrile illness, followed by hemorrhagic fever, encephalitis, and ocular disease, and can lead to death [2]. It is endemic in Sub-Saharan Africa, being first isolated in Kenya in 1930 [3]. Outbreaks were limited to that region until 1977-1978, when the virus spread to Egypt [4]; in 1993, southern Egypt suffered another outbreak, in which 600-1500 human infections were reported [5]. Periodic RVFV epizootics and epidemics have been associated with above-average rainfall and other environmental factors that result in dramatically increased mosquito populations [6,7].

A recurrence of Rift Valley Fever in East Africa was reported in 1997-1998 [8]. In 1987, a first West African epidemic occurred in Senegal and Mauritania during flooding in the lower Senegal River area [9]. The first outbreaks outside Africa occurred in 2000, in Saudi Arabia and Yemen [10]. In 2000-2010, outbreaks were reported in Sudan, Kenya, Tanzania, Somalia, Senegal, Mauritania, and Swaziland, with incidence rates higher than in the 1978 Egyptian epidemic [11-14]. RVFV has not apparently become endemic outside Africa, but seropositive animals have been detected in Saudi Arabia [15]. Climate conditions are appropriate for incursions of RVFV elsewhere in the Middle East, Europe, and beyond [10,16].

RVFV has been isolated from both livestock and mosquitoes [17,18]. The virus is maintained in mammal host species, including cattle, sheep, goats, and camels, in which infections have been reported [18,19]. RVFV is transmitted via several routes: mosquitoes serve as vectors in most cases, but direct transmission through aerosol and contact with abortion

products are other routes [20]. RVFV is also capable of persisting in the environment for long periods between epidemics [21], facilitated by vertical transmission among mosquitoes [22].

The RVFV genome is organized in three negative-sense, single-stranded RNA segments termed large (L), medium (M), and small (S), with a total genome length of 11.9 kb. The large segment (~6.4 kb) encodes the RNA-dependent RNA polymerase [23]; the M segment (~3.2 kb) encodes envelope glycoproteins G_n and G_c, plus two accessory proteins, NSm and the 78-kDa protein [24]. The S ambisense segment (~1.7 kb) encodes for nucleoprotein (NP; 27 kDa) and non-structural protein (NSs; 31-kDa). Previous studies have sequenced the three virus segments from diverse strains circulating in outbreaks across Africa and Saudi Arabia [25].

Historical movements of RVFV among countries raise concerns about possible appearance of RVFV in new regions [16]. Here, we aim to derive a detailed picture of RVFV phylogeny based on analysis of sequences of the three segments. We used phylogeographic approaches to examine mobility patterns of virus lineages across the virus' geographic distribution and among hosts.

Materials and Methods

Data used in this analysis represent all RVFV strains deposited in GenBank (as of August 2014; http://www.ncbi.nlm.nih.gov/nuccore), and include full sequences of the L, M, and S segments. If two or more records were available from the same isolate, we included the more recently sequenced version in analyses. Sequences for which the GenBank metadata listed no country of origin were excluded from those analyses; similarly, sequences lacking data on host were excluded from analyses of host associations. Sequences were aligned using the MUSCLE plugin [26] in the MEGA 6 software [27].

We used JModelTest [28] to identify the best-fitting nucleotide substitution model for each of the segments separately. A molecular clock-based phylogenetic analysis was performed for each segment separately in BEAST [29], using the best-fitting nucleotide substitution model for each segment, an uncorrelated lognormal relaxed molecular clock [30], and a GMRF Bayesian skyride tree prior [31]. Because variation among virus sampling dates (i.e., 1944-2010) is of meaningful amplitude relative to the time to most recent common ancestor of the clade in question [32,33], the temporal information associated with sampling each isolate had to be taken into consideration [30,34]. Sampling dates were used as prior information to calibrate the tree, estimate ages of different RVFV lineages, and infer evolutionary history of sampled strains [29]. An uncorrelated lognormal relaxed molecular clock was used in light of its high accuracy and precision to infer temporal information into molecular phylogeny [30]. Codon positions 1 and 2, and codon position 3, were treated as two separate partitions in the alignment. Multiple Monte Carlo Markov Chain (MCMC) runs of 10^8 states (the first 10% was discarded as burn-in) were combined to achieve estimated sample sizes of at least 250 for all numerical model parameters.

The posterior set of trees from each of the three initial BEAST analyses was used as an empirical tree set for a discrete-trait phylogeography analysis [35]. We assumed an asymmetrical rate matrix. For each tree sampled from the MCMC, Markov Jumps procedure [36] was used to reconstruct a stochastic realization of the between-country diffusion process; results were summarized over the entire posterior distribution by calculating median numbers of transitions between each pair of countries and the posterior probability that at least one transition occurred.

BEAST was re-run on data with host information, with the same configuration as before, to generate further empirical tree sets. Sequences were available from samples of human, bovine, mosquito, ovine, and caprine origins. We chose to include mosquito sequences in light of

evidence of persistence of viral lineages via vertical transmission; in this sense, mosquitoes can be regarded as potential hosts in their own right, rather than simply as transmission vectors [22]. Markov Jumps was again used to reconstruct transitions between host species, as in the phylogeographic analysis, again assuming an asymmetric rate matrix.

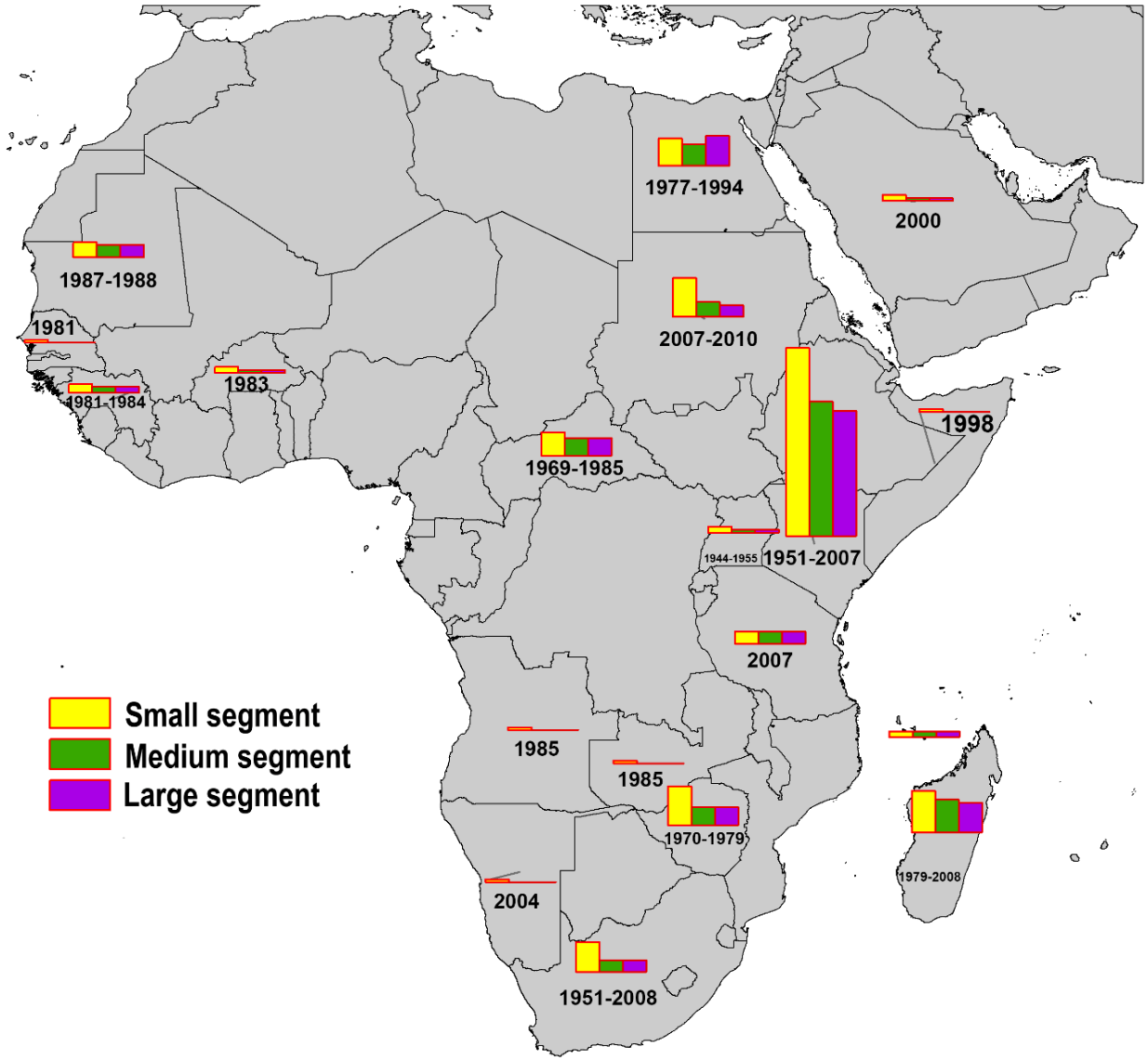


Fig 1: A map of countries with Rift Valley Fever outbreaks, with intensity of sequence sampling across Sub-Saharan Africa and the Arabian Peninsula. The bars on the map show numbers of sequences available for each minigenomic segment of RVFV. Dates in each country represent the range of years of origin of sequences identified from each country.

Results

Sequence data

A total of 155 S, 99 M, and 97 L minigenomic segments of RVFV sequence data were available on Genbank (Fig 1). Sequences had lengths of 1689-1692 bp for S, 3871-3885 bp for M, and 6397-6404 bp for L. These sequences represented RVFV strains from 18 countries across Africa plus Saudi Arabia (Fig 1). Saudi Arabian strains were represented by sequences for two S segments, one M segment, and one L segment only. Full details of the sequence data are available via Figshare (DOI: 10.6084/m9.figshare.2198776), including information for sequence accession, sequence length, country of origin, and isolation host.

Molecular clock and skyride analysis of RVFV strains

The nucleotide substitution models with the lowest Akaike Information Criterion scores identified by JModelTest were TPM2uf+I+G for the S segment, and GTR+I+G for both M and L. The maximum clade credibility (MCC) tree for the M segment is presented in Fig 2; trees for the other two segments can be found in the supplementary materials (S1-2 File). Lineages previously identified and discussed by Bird et al. [32] are indicated on the trees.

Estimated posterior mean nucleotide substitution rates were 3.6392×10^{-4}, with 95% highest posterior density (HPD) intervals of 2.8114×10^{-4} to 4.5813×10^{-4} substitutions per site per year for the S segment, 3.7774×10^{-4} (2.7391×10^{-4} to 4.8902×10^{-4}) for M, and 2.731×10^{-4} (1.9289×10^{-4} to 3.6677×10^{-4}) for L. The posterior mean time of the most recent common ancestor (TMRCA) of all isolates was 1929 (1920-1937) for S, 1914 (1897-1928) for M, and 1909 (1888 -1927) for L.

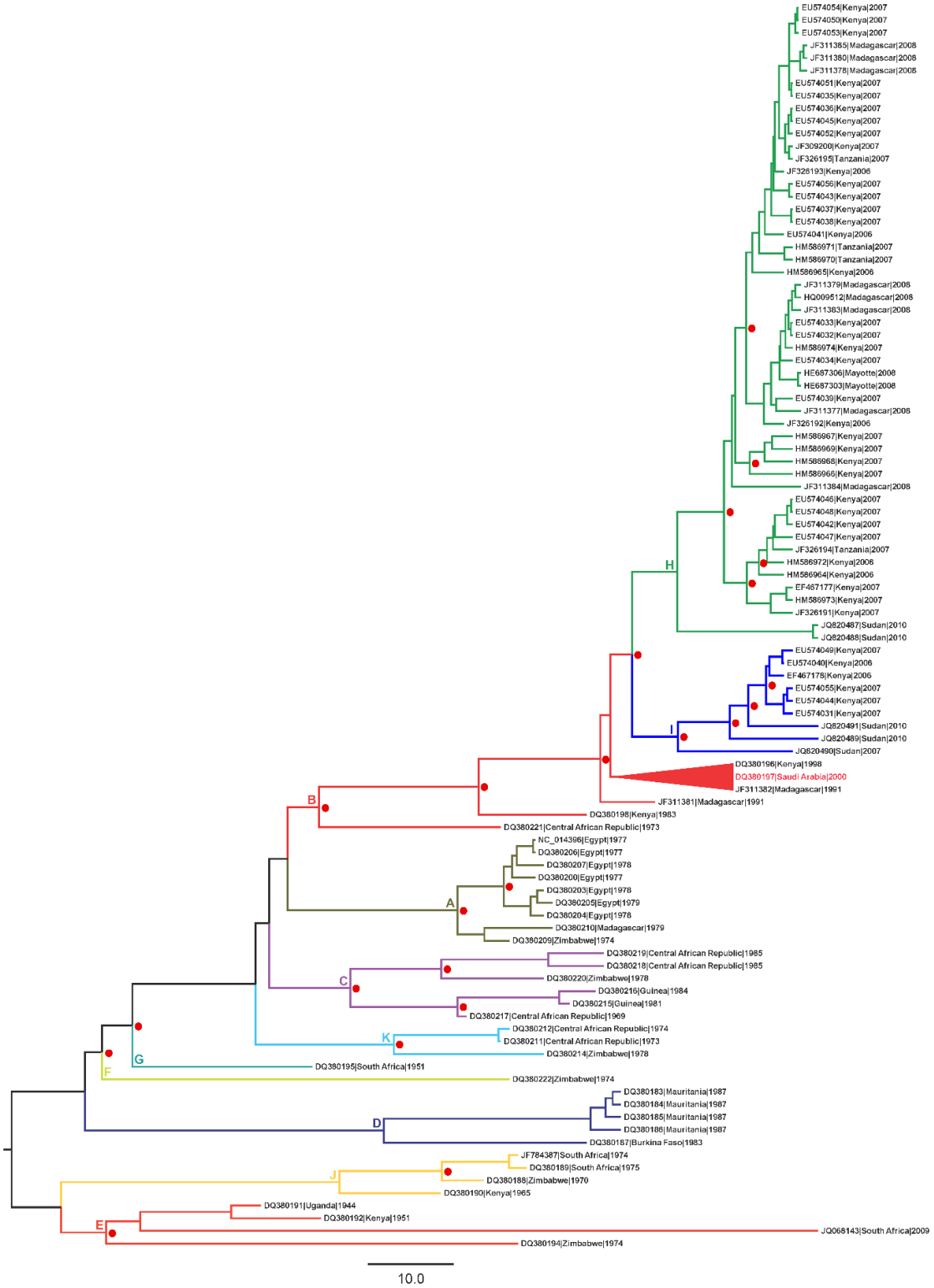


Fig 2: Maximum Clade Credibility tree based on all sequences of the medium minigenomic segment (M) of RVFV isolates in the study. Accession number, country, and date of sampling are presented at the tree tips. Tree branches are colored and labelled alphabetically by lineage (A to K). Lineage nomenclature is from Bird et al. [32]. Red triangle identifies the relationship of RVFV from Saudi Arabia to that from Africa. Clades with posterior probability >0.9 are labelled with red circles.

Skyride plots reconstructing temporal variation in RVFV genetic diversity are presented in Fig 3 for all three segments. All three indicate a peak in diversity around the middle of the twentieth century, followed by a decline and levelling off, with a subsequent increase in the reconstruction for the S segment.

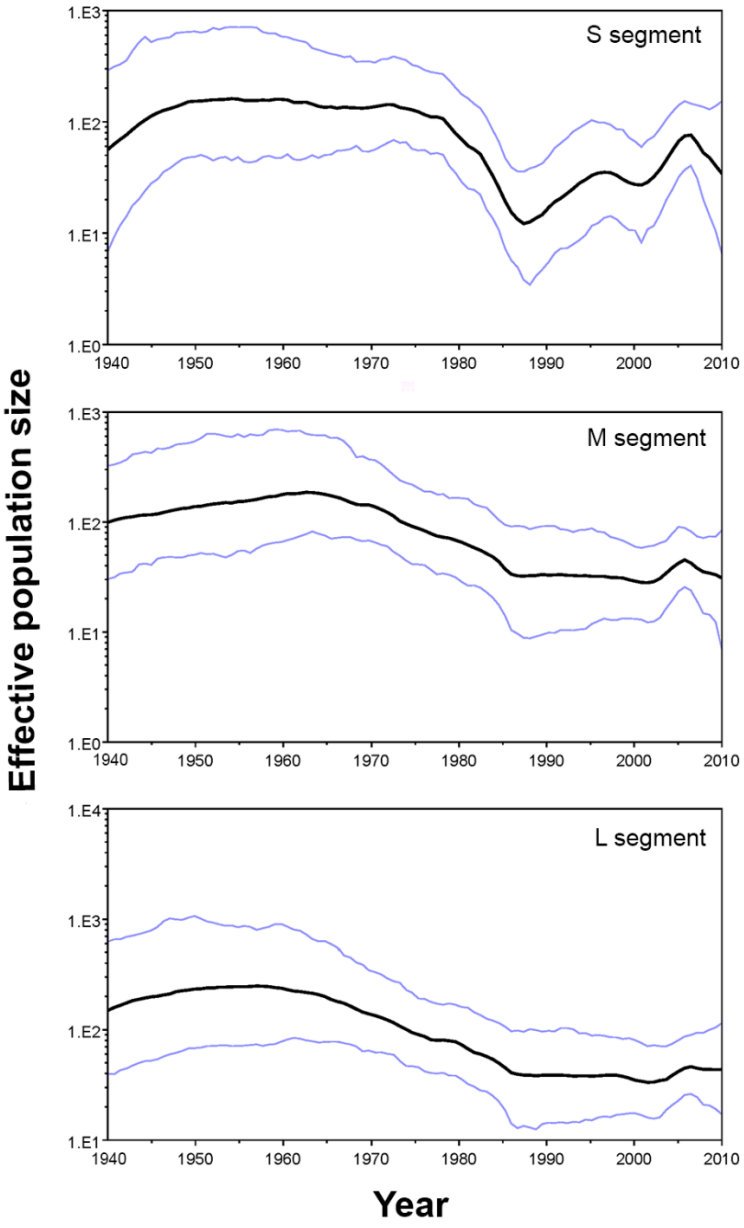


Fig 3: Gaussian Markov Random Field (GMRF) Bayesian skyride plot, representing the relationship between effective population size and time in years. Blue lines show the boundaries of the 95% highest posterior density interval.

Phylogeography of RVFV strains

Fig 4 presents the MCC phylogeny for the M segment, this time with branches colored by highest posterior probability of location; trees for the other two segments are in the supplementary materials (S 3-4 File). The tree shows that strains from Saudi Arabia belong to the same lineage (B) as those from Kenya in 1998 and Madagascar in 1991. Most strains from West Africa (Mauritania and Burkina Faso) are part of lineage D, but those from Guinea are in lineage C, which is otherwise recorded in the Central African Republic and Zimbabwe.

The diffusion patterns for viral lineages reconstructed using Markov Jumps are presented in Fig 5 & S 5-6 File; The complete results of the Markov Jumps analysis for all countries are available via Figshare (DOI: 10.6084/m9.figshare.2198776). The reconstruction using all three segments revealed that the highest median number of jumps was from Kenya to other countries in East Africa; movements of RVFV lineages from Kenya to Tanzania are particularly well supported (posterior probability >0.9). Hence, arrival of the virus in Tanzania in 2007 was probably related to a single introduction event from Kenya. Introductions of RVFV to Madagascar came in three waves: the first from Zimbabwe, and the second and third from Kenya.

The two major outbreaks in Egypt since 1977 originally may have been the result of a long-distance introduction from Zimbabwe, as they are closely related to strains from Zimbabwe in 1974. Markov Jumps revealed a possible transition from Kenya to Mayotte, with posterior probabilities ranging from 0.83 to 0.96 in different genomic segments. All sequences from Sudan came from a single outbreak in 2007-2010, and were closely related to isolates from outbreaks in other East African countries since 2007; Markov Jumps suggested direct movement of the virus from Kenya to Sudan (posterior probability >0.7).

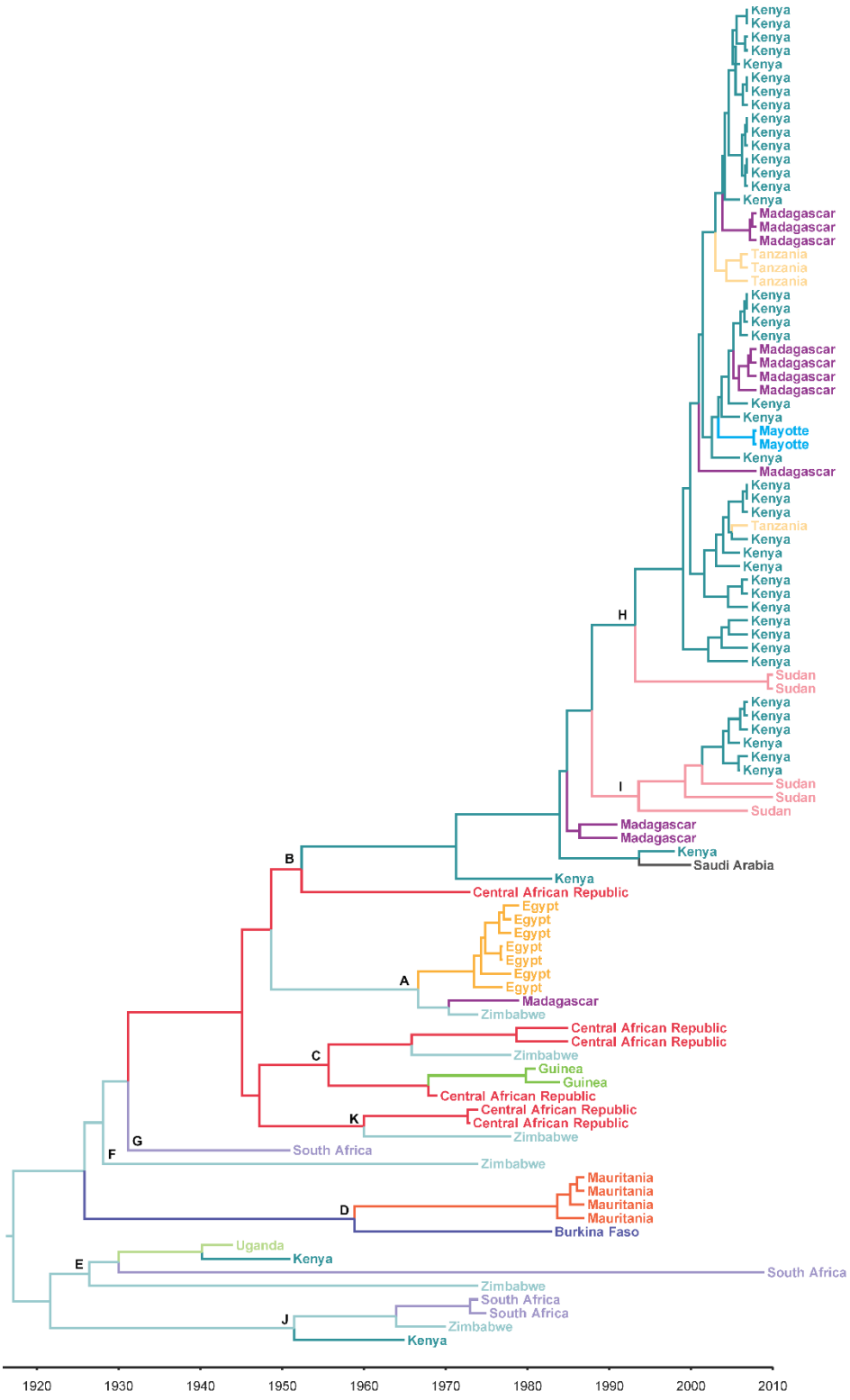


Fig 4: Maximum Clade Credibility tree based on all sequences of the medium minigenomic segment (M) of RVFV isolates in the study. Country of origin is indicated by color of the tree branches and branch tips.

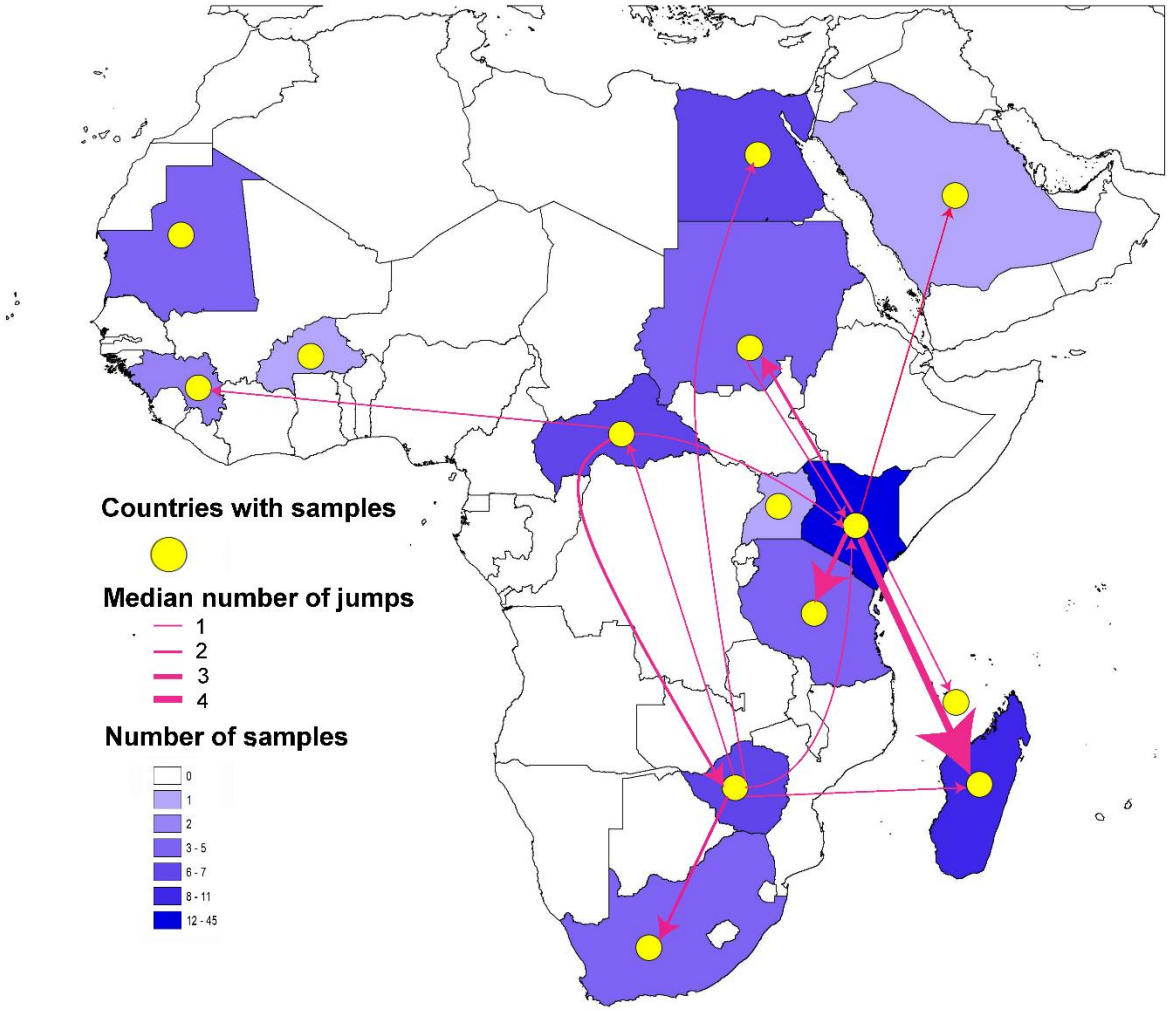


Fig 5: Connectedness of countries with Rift Valley Fever outbreaks based on Markov Jumps analysis of the medium minigenomic segment (M) of RVFV isolates in the study. The map shows only countries with non-zero transition frequencies. Connections between countries are presented as lines with arrows to refer to the direction of movement. Line thickness identifies the median number of jumps between each country pair.

Markov Jumps analysis of M and L segments indicated virus introduction from Zimbabwe to the Central African Republic (posterior probabilities 0.51 and 0.64, respectively). Other movement patterns inferred included transitions from Zimbabwe to South Africa (posterior probabilities 0.59, 0.68, and 0.78 for S, M, and L, respectively). M and L segments revealed a single transition into West Africa, from the Central African Republic to Guinea.

The M and L segments suggested a single introduction from Kenya to Saudi Arabia, with a posterior probability >0.79 for both segments. Movement of the virus between Kenya and Sudan occurred in both directions, with posterior probabilities of >0.7. Similarly, transitions between Central African Republic and Zimbabwe were reconstructed in both directions (posterior probabilities 0.57, 0.79, and 0.69 for S, M, and L, respectively, for Central African Republic to Zimbabwe; posterior probabilities 0.62, 0.51, and 0.64 for Zimbabwe to Central African Republic, for S, M, and L, respectively). The S segment revealed information about movements involving countries for which it was the only segment available: from Kenya to Somalia (posterior probability 0.77), Burkina Faso to Mauritania (posterior probability 0.80), Burkina Faso to Senegal (posterior probability 0.78), Kenya to Central African Republic (posterior probability 0.91), and Kenya to South Africa (posterior probability 0.91) (S File 6).

RVFV host transition analysis

GenBank records for which host data were available numbered 153 for the S segment, 98 for M, and 94 for L. The MCC tree for the M segment with branches colored by the highest posterior probability host species is shown in Fig 6; trees for other segments are in the supplementary materials (S 7-8 File). Humans and cattle were the most likely hosts of the most recent common ancestor (MRCA) of all isolates, with posterior probabilities of 0.377 and 0.350, respectively, as analysis suggested circulation of the virus in both humans and bovines since the 1920s. The L segment had probabilities of 0.500 for humans and 0.233 for cattle. Finally, the S segment suggested that the host of the MRCA was a mosquito (posterior probability 0.796), but this outcome is most likely biased by the fact that three of the earliest five sequences were isolated from mosquitoes, which is not the case in the other two segments.

A summary of the Markov Jumps results for host associations is presented in Table 1. This analysis revealed ≥95% support for two types of host transitions: from human to bovine and bovine to human. The S and L segments also showed ≥95% support for a human to mosquito transition, and the S segment alone indicated a mosquito to human, with ≥95% support. No other transition type had strong support.

	No. of jumps to destination					
Origin	Human	Bovine	Ovine	Caprine	Mosquito	Bat
Human		13 (1.00)	3 (0.94)	3 (0.97)	10 (1.00)	0 (0.11)
		14 (1.00)	3 (0.93)	1 (0.50)	8 (0.99)	0 (0.10)
		15 (1.00)	3 (0.93)	1 (0.68)	7 (0.99)	1 (0.80)
Bovine	6 (0.99)		1 (0.78)	0 (0.30)	2 (0.87)	0 (0.07)
	6 (0.99)		2 (0.88)	1 (0.93)	2 (0.87)	0 (0.08)
	5 (0.98)		1 (0.69)	1 (0.63)	1 (0.50)	0 (0.13)
Ovine	0 (0.18)	0 (0.18)		0 (0.24)	0 (0.16)	0 (0.03)
	0 (0.39)	0 (0.30)		0 (0.02)	0 (0.30)	0 (0.04)
	0 (0.27)	0 (0.28)		0 (0.06)	0 (0.32)	0 (0.04)
Caprine	0 (0.07)	0 (0.04)	1 (0.52)		0 (0.03)	0 (0.01)
	0 (0.03)	0 (0.14)	0 (0.03)		0 (0.04)	0 (0.02)
	0 (0.04)	0 (0.08)	0 (0.08)		0 (0.03)	0 (0.02)
Mosquito	5 (0.99)	3 (0.85)	2 (0.88)	0 (0.12)		1 (0.81)
	1 (0.76)	1 (0.53)	1 (0.67)	0 (0.10)		1 (0.81)
	1 (0.65)	0 (0.31)	1 (0.59)	0 (0.15)		0 (0.07)
Bat	0 (0.01)	0 (0.02)	0 (0.02)	0 (0.00)	0 (0.03)	
	0 (0.01)	0 (0.03)	0 (0.02)	0 (0.00)	0 (0.05)	
	0 (0.03)	0 (0.03)	0 (0.06)	0 (0.00)	0 (0.02)	

Table 1: Median numbers of reconstructed Markov jumps between host pairs in the host transition analyses, presented as median values across all trees in the posterior sample. Results from all segments are presented in each cell: S (top), M (middle), and L (bottom). Numbers in parenthesis are the posterior probability that at least one jump occurred since the time of the common ancestor of all sequences in each minigenomic segment.

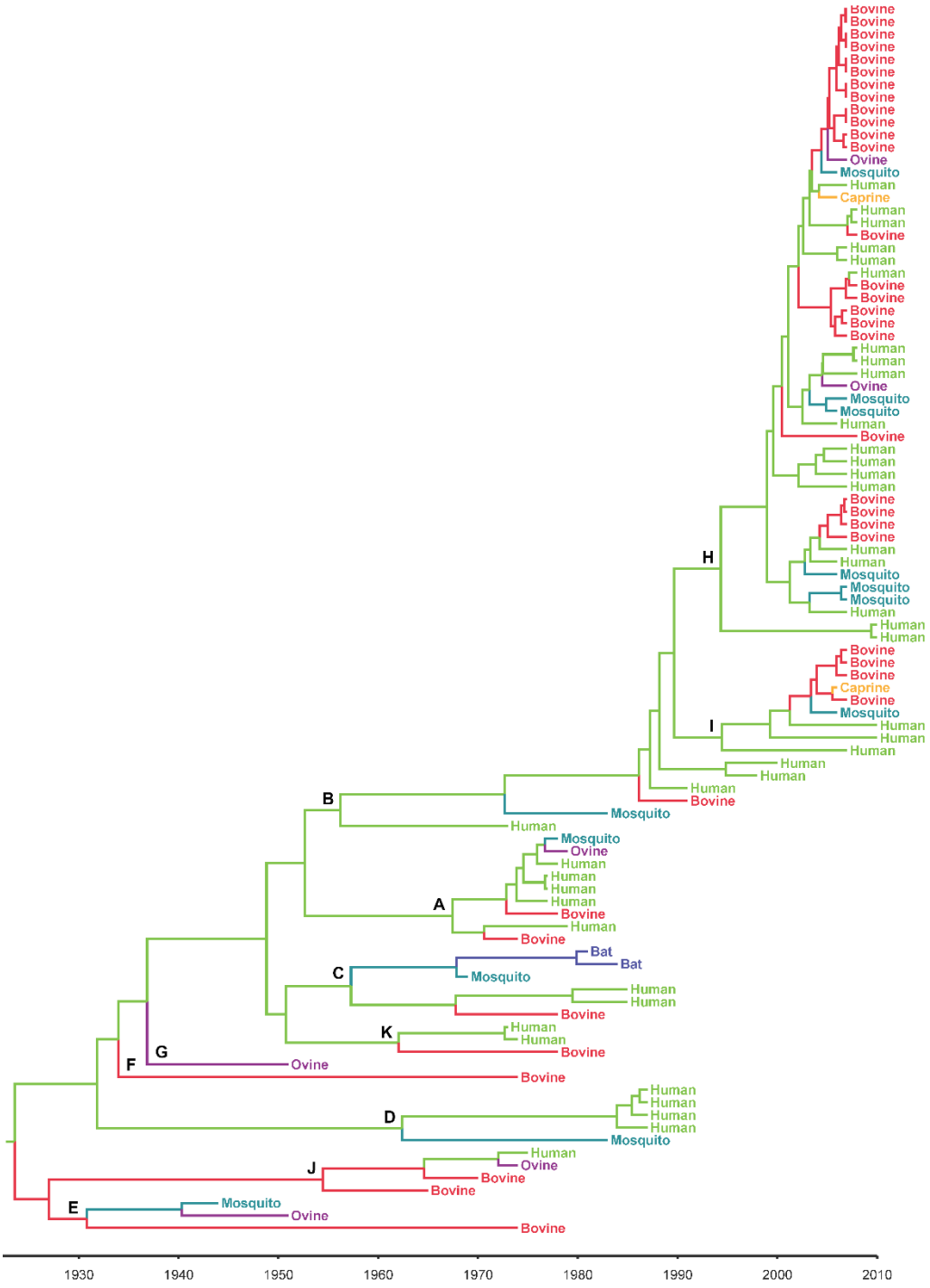


Fig 6: Maximum Clade Credibility tree from all sequences of the medium minigenomic segment (M) of RVFV isolates in the study. The color of branches and branch tips identifies the host of each RVFV strain.

Discussion

This study used novel phylogenetic approaches [35] to investigate the ancestry of RVFV strains across Africa and Saudi Arabia, and to study virus movements and host transitions. Sequences for the three segments of the RVFV genome were available for strains sampled over a span of 66 years (1944-2010).

Previous RVFV studies have revealed no evidence of recombination in RVFV [37,38], so it should not be a cause for concern in this analysis [37,38]. On the other hand, evidence exists for reassortment among RVFV segments [39]: Freire et al. [33] identified seven reassortment events among segments. RVFV minigenomic segments showed differences in amounts of genetic change and time scale [33]. These minigenomic segments have previously been used to derive phylogenetic and ecological insights regarding RVFV circulation in Africa and the Arabian Peninsula [32,33]. These analyses used phylogenetic approaches similar to ours; however, our analysis takes advantage of a discrete-traits phylogenetic analysis using Markov Jumps to infer the history of between-country and between-host movements.

The substitution rate estimates in our study were similar to those in previous studies [33,40] . Other studies have reported higher rates [32,41]: for example, estimates from Aradaib *et al*. [41] were 4.20×10^{-4}, 5.06×10^{-4}, and 4.29×10^{-4} substitutions per site per year for S, M, and L, respectively; the 95% HPD intervals reported in that paper overlap with ours. These differences might be a result of the different datasets used: our work should have better resolution because the dataset is larger and more diverse. The Aradaib *et al*. [41] study was also limited to only two lineages of the 11 in our analysis, and covered only strains from Kenya, Sudan, Madagascar, and Zimbabwe.

Previous studies reported earlier TMRCA estimates than ours. Bird *et al*. 2007 [32] estimated the mean TMRCA as 1891 for the S segment, 1882 for M, and 1887 for L. These differences could be a result of our larger dataset, and reflect higher estimates of substitution rates; again, their HPD estimates overlap with ours. As the HPD intervals for the TMRCAs of the three segments in our analysis all overlap, the difference in point estimates likely reflects statistical uncertainty only. Our TMRCA estimates for all segments agreed well with the first report of RVFV in 1930 in Kenya [3].

The steady decline in RVFV genetic diversity since the 1970's was previously reported [33], in an analysis that considered most of our samples. A possible explanation for the decline centers on the vaccination and control measures implemented on a large scale from 1969 to 1979 [42]. In all, 35.2 million vaccines were provided to Zimbabwe, South Africa, Namibia, Israel, and Egypt in response to large RVF outbreaks [43].

Discrete-traits phylogenetic approaches have some limitations [34]. In our case, these limitations are associated with the nature of virus sampling across its range, as sampling is generally unbalanced. This point suggests that some aspects of our results should be interpreted with caution. For example, in our host analyses, most samples came from cattle, sheep, goats, and humans; however, sampling of bats was limited to two isolates from Guinea, probably insufficient to conclude that bats are not historically associated with RVFV. Country was used as indication of location in our analyses, which is quite coarse for some of the spatial phenomena that we would like to reconstruct. As more sequence data become available, it should be possible to develop finer-resolution views.

In Sub-Saharan Africa, RVFV appears to be spread by movement of viremic livestock between countries [44,45], or though introduction of infected mosquitoes to neighboring

countries [16,45]. We used Markov Jumps to infer possible introduction events and movement routes of RVFV. Two types of RVFV movements can be considered: short- and long-distance jumps [32,46,47]. Inferred movements between distant countries may omit the effects of unsampled lineages in countries on the route between them. For example, previous studies attributed the 1977-1978 epidemics in Egypt to viral introductions from Sudan [48,49], but our study saw strong support for Zimbabwe as a country of origin for Egyptian strains. This result suggests that, although Zimbabwe was the sampled origin for these lineages, they travelled north over Sudan to Egypt; all available Sudanese sequences came from more recent outbreaks in 2007-2010, which presumably originated in Kenya, and are genetically distant from Egyptian strains in the outbreak of the 1970s [41]. With no earlier Sudanese sequences available [50,51], this analysis could not find an origin in Sudan, and hence tracked lineages back to Zimbabwe; this result should thus not be taken to indicate that the hypothesis of a Sudanese origin for Egyptian epidemics is incorrect.

The RVFV strain identified from the Arabian Peninsula in 2000 was embedded in lineage B with strains from Kenya, suggesting that this virus originated from Kenyan epizootics in 1997-1998. The outbreak was driven by floods and heavy rains along the Saudi Arabia-Yemen border in the Al Humayrha region, where the first cases were reported [52], and where it was maintained by *Culex tritaeniorhynchus* [53].

RVFV strains from West Africa fell in two lineages (C & D): one included samples from Guinea and another that included samples from Burkina Faso, Mauritania, and Senegal. Our results suggest possible introduction of RVFV to Guinea from the Central African Republic, and that the outbreak in Mauritania in 1987 had its origin in the event in lineages that were in Burkina Faso in 1983. The route of introduction from East Africa to West Africa more

fundamentally is still unclear. Our analysis suggested interesting patterns for outbreaks in Mauritania and Egypt, in comparison to recent outbreaks in Kenya, with single viral introductions to Mauritania and Egypt, but multiple origins for the 2007-2010 outbreaks in Kenya.

The final analysis presented in this paper provides a coarse-resolution picture of major tendencies in host transitions. RVFV has been isolated from several livestock and wild mammal species, as well as humans [2]; bats and rodents have also been suspected as reservoirs of the virus [54]. Nevertheless, the role of wild animals in the persistence and emergence of RVFV remains obscure, with studies in Madagascar revealing absence in rodents [54], and no clear evidence for a role of bats in maintenance of RVFV populations [55]. RVFV has been isolated from livestock across Africa since its first isolation [3]; however, association of wild animals remain unclear and experimental infection has revealed low concentration of RVFV antigens [55]. Our results were consistent with major roles of bovines, humans, and mosquitoes in transmission across all outbreaks in Africa [2,56]; however, the results here should be interpreted with caution owing to the incomplete nature of the sampling.

RVFV may circulate in a wider range of hosts that have not been investigated in previous studies, and hence have provided no sequence data. For example, birds have been found to be refractory for RVFV and may be possible host candidates; at present, however, no genetic data are available that would allow us to include them in an analysis of this sort. Also persistent vertical transmission of RVFV among mosquitoes makes them a possible long-term host species; however, the data in this study are obviously insufficient to capture every possible lineage that may have been present in a mosquito, and any transition we infer between mammalian species generally had mosquitoes as an intermediary. Several other studies have outlined the difficulties

in detecting the virus in mosquitoes, especially during inter-epidemic periods [56]. The unbalanced nature of the dataset with respect to host species marks a limitation for our analysis. Future studies should collect samples more systematically, and on a much finer scale with respect to location and host to give a more detailed picture of migratory patterns of RVFV across the continent.

Acknowledgments

AMS was supported by the Egyptian Fulbright Mission Program. The authors would like to thank the Department of Entomology and the Research and Training Center on Vectors of Diseases of Ain Shams University, Egypt for their continuous support. Special thanks to colleagues in the KU ENM Group for their support during the study.

References

1. Balkhy HH, Memish ZA (2003) Rift Valley Fever: An uninvited zoonosis in the Arabian Peninsula. Int J Antimicrob Agents. 21: 153-157.
2. Ikegami T, Makino S (2011) The pathogenesis of Rift Valley Fever. Viruses 3: 493-519.
3. Daubney R, Hudson JR, Garnham PC (1931) Enzootic hepatitis or Rift Valley Fever: An undescribed virus disease of sheep cattle and man from East Africa. J Pathol Bacteriol. 34: 545-579.
4. Ahmed Kamal S (2011) Observations on Rift Valley Fever virus and vaccines in Egypt. Virol J. 8: 532.

5. Arthur RR, el-Sharkawy MS, Cope SE, Botros BA, Oun S, et al. (1993) Recurrence of Rift Valley Fever in Egypt. Lancet 342: 1149-1150.

6. Metras R, Jewell C, Porphyre T, Thompson PN, Pfeiffer DU, et al. (2015) Risk factors associated with Rift Valley Fever epidemics in South Africa in 2008-11. Sci Rep. 5: 9492.

7. Caminade C, Ndione JA, Diallo M, MacLeod DA, Faye O, et al. (2014) Rift Valley Fever outbreaks in Mauritania and related environmental conditions. Int J Environ Res Public Health. 11: 903-918.

8. Lichoti JK, Kihara A, Oriko AA, Okutoyi LA, Wauna JO, et al. (2014) Detection of Rift Valley Fever virus interepidemic activity in some hotspot areas of Kenya by sentinel animal surveillance, 2009-2012. Vet Med Int. 2014: 379010

9. Digoutte JP, Peters CJ (1989) General aspects of the 1987 Rift Valley Fever epidemic in Mauritania. Res Virol. 140: 27-30.

10. Madani TA, Al-Mazrou YY, Al-Jeffri MH, Mishkhas AA, Al-Rabeah AM, et al. (2003) Rift Valley Fever epidemic in Saudi Arabia: Epidemiological, clinical, and laboratory characteristics. Clin Infect Dis. 37: 1084-1092.

11. Jost CC, Nzietchueng S, Kihu S, Bett B, Njogu G, et al. (2010) Epidemiological assessment of the Rift Valley Fever outbreak in Kenya and Tanzania in 2006 and 2007. Am J Trop Med Hyg. 83: 65-72.

12. Himeidan YE, Kweka EJ, Mahgoub MM, El Rayah el A, Ouma JO (2014) Recent outbreaks of Rift Valley Fever in East Africa and the Middle East. Front Public Health. 2: 169.

13. Chevalier V, Thiongane Y, Lancelot R (2009) Endemic transmission of Rift Valley Fever in Senegal. Transbound Emerg Dis. 56: 372-374.

14. Glancey MM, Anyamba A, Linthicum KJ (2015) Epidemiologic and environmental risk factors of Rift Valley Fever in southern Africa from 2008 to 2011. Vector Borne Zoonotic Dis. 15: 502-511.

15. Mohamed AM, Ashshi AM, Asghar AH, Abd El-Rahim IH, El-Shemi AG, et al. (2014) Seroepidemiological survey on Rift Valley fever among small ruminants and their close human contacts in Makkah, Saudi Arabia, in 2011. Rev Sci Tech. 33: 903-915.

16. Chevalier V, Pepin M, Plee L, Lancelot R (2010) Rift Valley Fever--A threat for Europe? Euro Surveill. 15: 19506.

17. Munyua P, Murithi RM, Wainwright S, Githinji J, Hightower A, et al. (2010) Rift Valley fever outbreak in livestock in Kenya, 2006-2007. Am J Trop Med Hyg. 83: 58-64.

18. Sumaye RD, Geubbels E, Mbeyela E, Berkvens D (2013) Inter-epidemic transmission of Rift Valley fever in livestock in the Kilombero River Valley, Tanzania: A cross-sectional survey. PLoS Negl Trop Dis. 7: e2356.

19. El Mamy AB, Lo MM, Thiongane Y, Diop M, Isselmou K, et al. (2014) Comprehensive phylogenetic reconstructions of Rift Valley Fever virus: the 2010 northern Mauritania outbreak in the *Camelus dromedarius* species. Vector Borne Zoonotic Dis. 14: 856-861.

20. Zeller HG, Fontenille D, Traore-Lamizana M, Thiongane Y, Digoutte JP (1997) Enzootic activity of Rift Valley Fever virus in Senegal. Am J Trop Med Hyg. 56: 265-272.

21. Craig DE, Thomas WJ, DeSanctis AN (1967) Stability of Rift Valley Fever virus at 4 C. Appl Microbiol. 15: 446-447.

22. Linthicum KJ, Davies FG, Kairo A, Bailey CL (1985) Rift Valley Fever virus (family *Bunyaviridae*, genus *Phlebovirus*): Isolations from Diptera collected during an inter-epizootic period in Kenya. J Hyg. 95: 197-209.

23. Gauliard N, Billecocq A, Flick R, Bouloy M (2006) Rift Valley Fever virus noncoding regions of L, M and S segments regulate RNA synthesis. Virology. 351: 170-179.

24. Suzich JA, Kakach LT, Collett MS (1990) Expression strategy of a *phlebovirus*: biogenesis of proteins from the Rift Valley Fever virus M segment. J Virol. 64: 1549-1555.

25. Cêtre-Sossah C, Zeller H, Grandadam M, Caro V, Pettinelli F, et al. (2012) Genome analysis of Rift Valley Fever virus, Mayotte. Emerg Infect Dis. 18: 969-971.

26. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32: 1792-1797.

27. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 30: 2725-2729.

28. Posada D (2008) jModelTest: Phylogenetic model averaging. Mol Biol Evol. 25: 1253-1256.

29. Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol. 29: 1969-1973.

30. Drummond AJ, Ho SY, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. PLoS Biol. 4: e88.

31. Minin VN, Bloomquist EW, Suchard MA (2008) Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. Mol Biol Evol. 25: 1459-1471.

32. Bird BH, Khristova ML, Rollin PE, Ksiazek TG, Nichol ST (2007) Complete genome analysis of 33 ecologically and biologically diverse Rift Valley Fever virus strains reveals widespread virus movement and low genetic diversity due to recent common ancestry. J Virol. 81: 2805-2816.

33. Freire CCM, Iamarino A, Soumaré POL, Faye O, Sall AA, et al. (2015) Reassortment and distinct evolutionary dynamics of Rift Valley Fever virus genomic segments. Sci Rep. 5: 11353.

34. Hall MD, Knowles NJ, Wadsworth J, Rambaut A, Woolhouse ME (2013) Reconstructing geographical movements and host species transitions of foot and mouth disease virus serotype SAT 2. mBio 4: e00591-00513.

35. Lemey P, Rambaut A, Drummond AJ, Suchard MA (2009) Bayesian phylogeography finds its roots. PLoS Comput Biol. 5: e1000520.

36. Minin VN, Suchard MA (2008) Counting labeled transitions in continuous-time Markov models of evolution. J Math Biol. 56: 391-412.

37. Chare ER, Gould EA, Holmes EC (2003) Phylogenetic analysis reveals a low rate of homologous recombination in negative-sense RNA viruses. J Gen Virol. 84: 2691-2703.

38. Grobbelaar AA, Weyer J, Leman PA, Kemp A, Paweska JT, et al. (2011) Molecular epidemiology of Rift Valley Fever virus. Emerg Infect Dis. 17: 2270-2276.

39. Sall AA, Zanotto PM, Sene OK, Zeller HG, Digoutte JP, et al. (1999) Genetic reassortment of Rift Valley Fever virus in nature. J Virol. 73: 8196-8200.

40. Pepin M, Bouloy M, Bird BH, Kemp A, Paweska J (2010) Rift Valley Fever virus (*Bunyaviridae*: *Phlebovirus*): an update on pathogenesis, molecular epidemiology, vectors, diagnostics and prevention. Vet Res. 41: 61.

41. Aradaib IE, Erickson BR, Elageb RM, Khristova ML, Carroll SA, et al. (2013) Rift Valley Fever, Sudan, 2007 and 2010. Emerg Infect Dis. 19: 246-253.

42. Kortekaas J, Zingeser J, de Leeuw P, de La Rocque S, Unger H, et al. (2011) Rift Valley Fever vaccine development: progress and constraints. Emerg Infect Dis. 17: e1.

43. Singh SK (2016) Human Emerging and Re-emerging Infections. John Wiley & Sons, Inc.

44. Nanyingi MO, Munyua P, Kiama SG, Muchemi GM, Thumbi SM, et al. (2015) A systematic review of Rift Valley Fever epidemiology, 1931–2014. Infect Ecol Epidemiol. 5: 28024.

45. Rolin AI, Berrang-Ford L, Kulkarni MA (2013) The risk of Rift Valley Fever virus introduction and establishment in the United States and European Union. Emerg Microbes Infect. 2: e81.

46. Metras R, Baguelin M, Edmunds WJ, Thompson PN, Kemp A, et al. (2013) Transmission potential of Rift Valley Fever virus over the course of the 2010 epidemic in South Africa. Emerg Infect Dis. 19: 916-924.

47. Balenghien T, Cardinale E, Chevalier V, Elissa N, Failloux AB, et al. (2013) Towards a better understanding of Rift Valley Fever epidemiology in the south-west of the Indian Ocean. Vet Res. 44: 78.

48. Gad AM, Feinsod FM, Allam IH, Eisa M, Hassan AN, et al. (1986) A possible route for the introduction of Rift Valley Fever virus into Egypt during 1977. J Trop Med Hyg. 89: 233-236.

49. Abdel-Wahab KS, El Baz LM, El-Tayeb EM, Omar H, Ossman MA, et al. (1978) Rift Valley Fever virus infections in Egypt: Pathological and virological findings in man. Trans R Soc Trop Med Hyg. 72: 392-396.

50. Eisa M (1984) Preliminary survey of domestic animals of the Sudan for precipitating antibodies to Rift Valley Fever virus. J Hyg. 93: 629-637.

51. Hassan OA, Ahlm C, Sang R, Evander M (2011) The 2007 Rift Valley fever outbreak in Sudan. PLoS Negl Trop Dis. 5: e1229.

52. Al-Afaleq AI, Hussein MF (2011) The status of Rift Valley Fever in animals in Saudi Arabia: A mini review. Vector Borne Zoonotic Dis. 11: 1513-1520.

53. Jupp PG, Kemp A, Grobbelaar A, Lema P, Burt FJ, et al. (2002) The 2000 epidemic of Rift Valley Fever in Saudi Arabia: Mosquito vector studies. Med Vet Entomol. 16: 245-252.

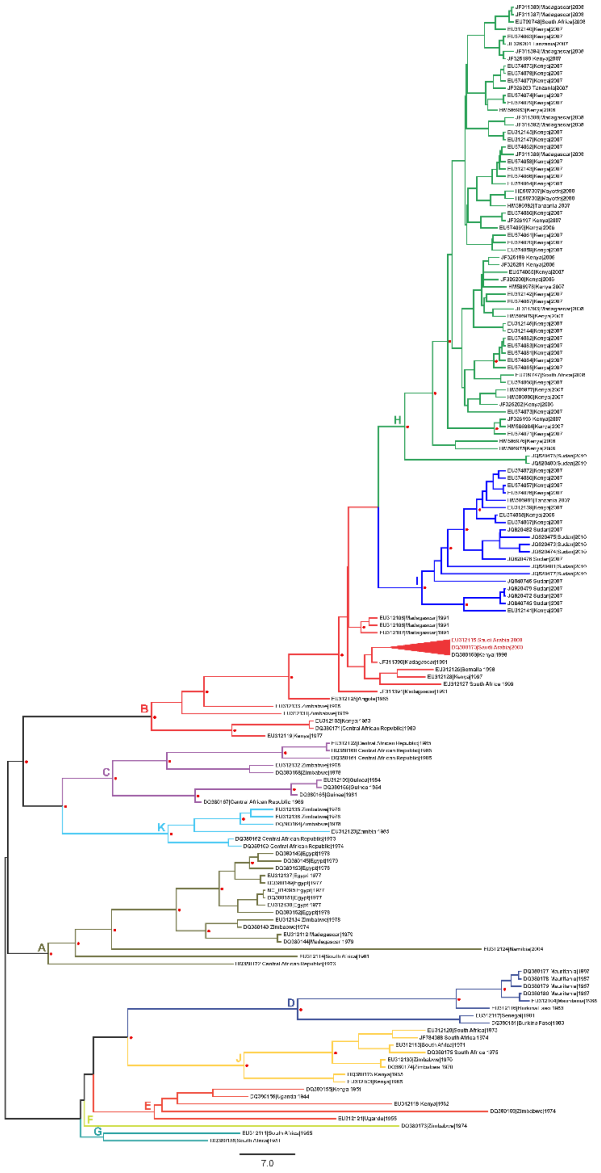
54. Olive MM, Goodman SM, Reynes JM (2012) The role of wild mammals in the maintenance of Rift Valley Fever virus. J Wildl Dis. 48: 241-266.

55. Oelofsen MJ, Van der Ryst E (1999) Could bats act as reservoir hosts for Rift Valley Fever virus? Onderstepoort J Vet Res. 66: 51-54.

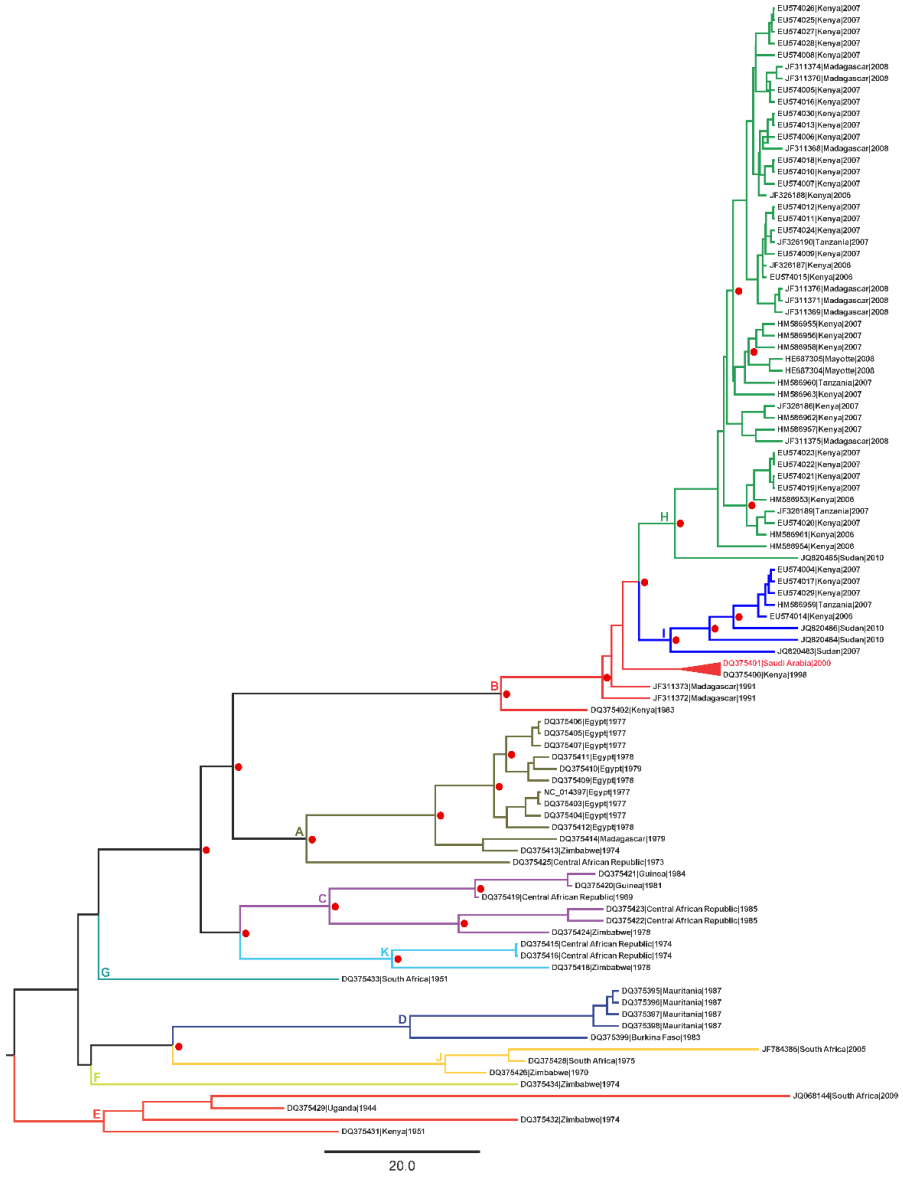
56. Mweya CN, Kimera SI, Mellau LS, Mboera LE (2015) Inter-epidemic abundance and distribution of potential mosquito vectors for Rift Valley Fever virus in Ngorongoro District, Tanzania. Glob Health Action. 8: 25929.

Supporting information

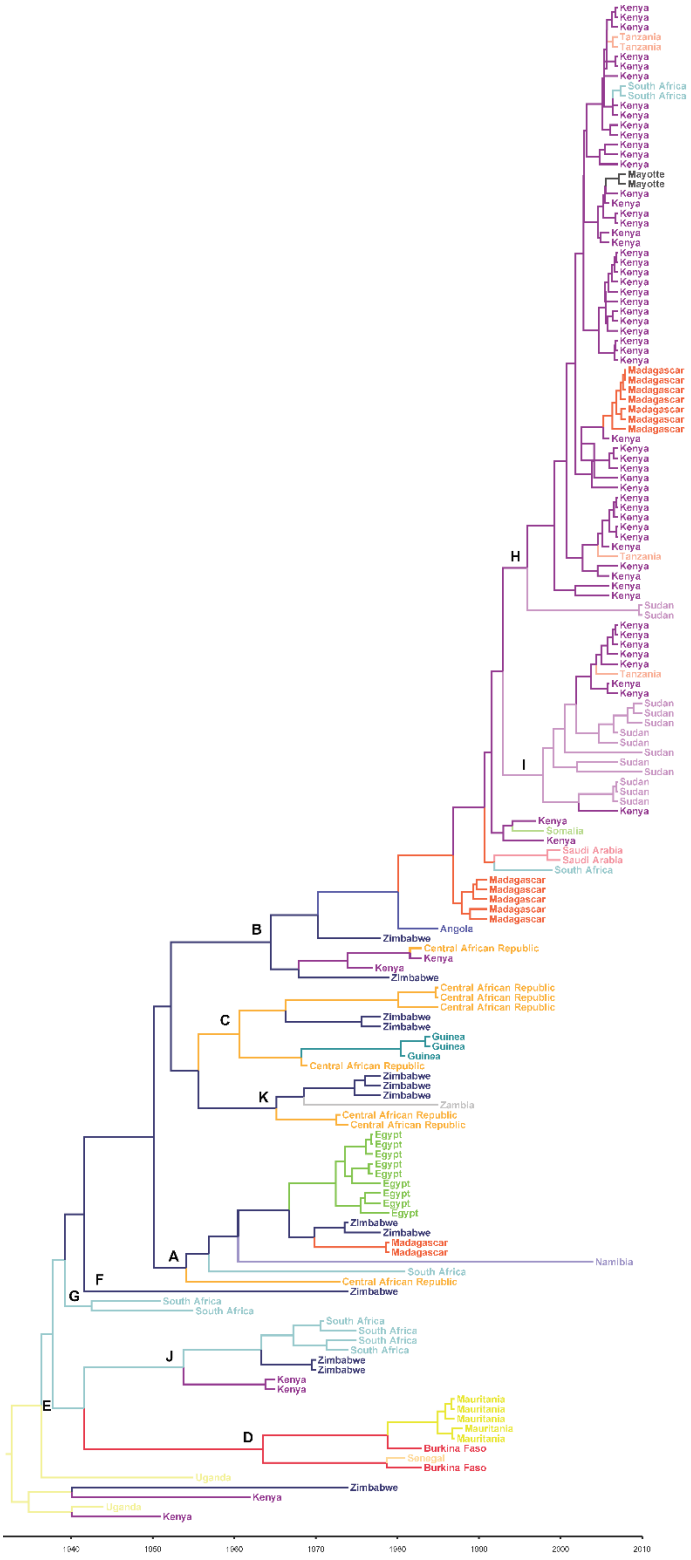
S1 File: Maximum Clade Credibility tree based on the small minigenomic segment (S) of RVFV isolates. Accession number, country, and date of sampling are presented at the tree tips. Tree branches are colored and labelled alphabetically by lineage (A to K). Lineage nomenclature is from Bird et al. [32]. Red triangle identifies the relationship of RVFV from Saudi Arabia to that from Africa. Clades with posterior probability >0.9 are labelled with red circles.



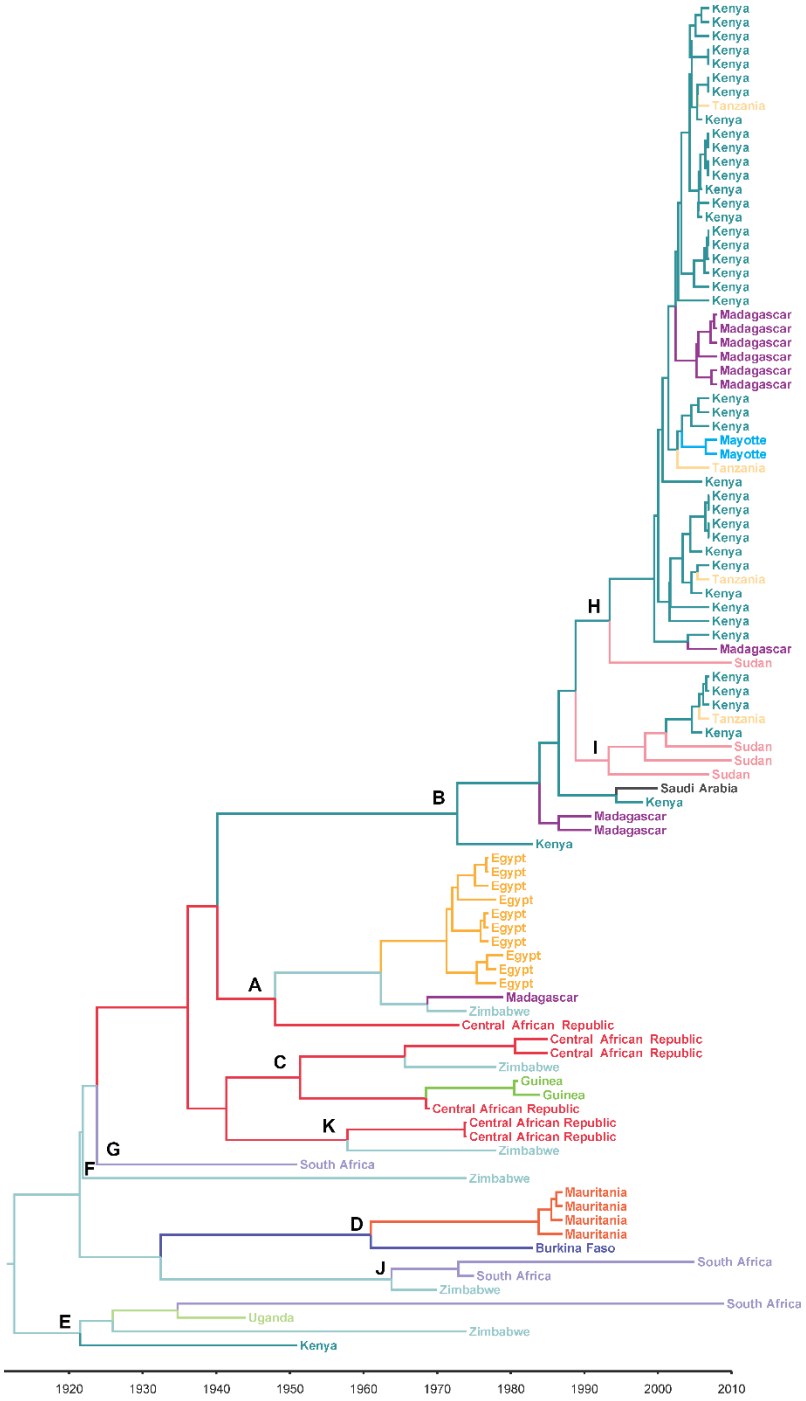
S2 File: Maximum Clade Credibility tree based on all sequences of the large minigenomic segments (L) of RVFV isolates in the study. Accession number, country, and date of sampling are presented at the tree tips. Tree branches are colored and labelled alphabetically by lineage (A to K). Lineage nomenclature is from Bird et al. [32]. Red triangle identifies the relationship of RVFV from Saudi Arabia to that from Africa. Clades with posterior probability >0.9 are labelled with red circles.



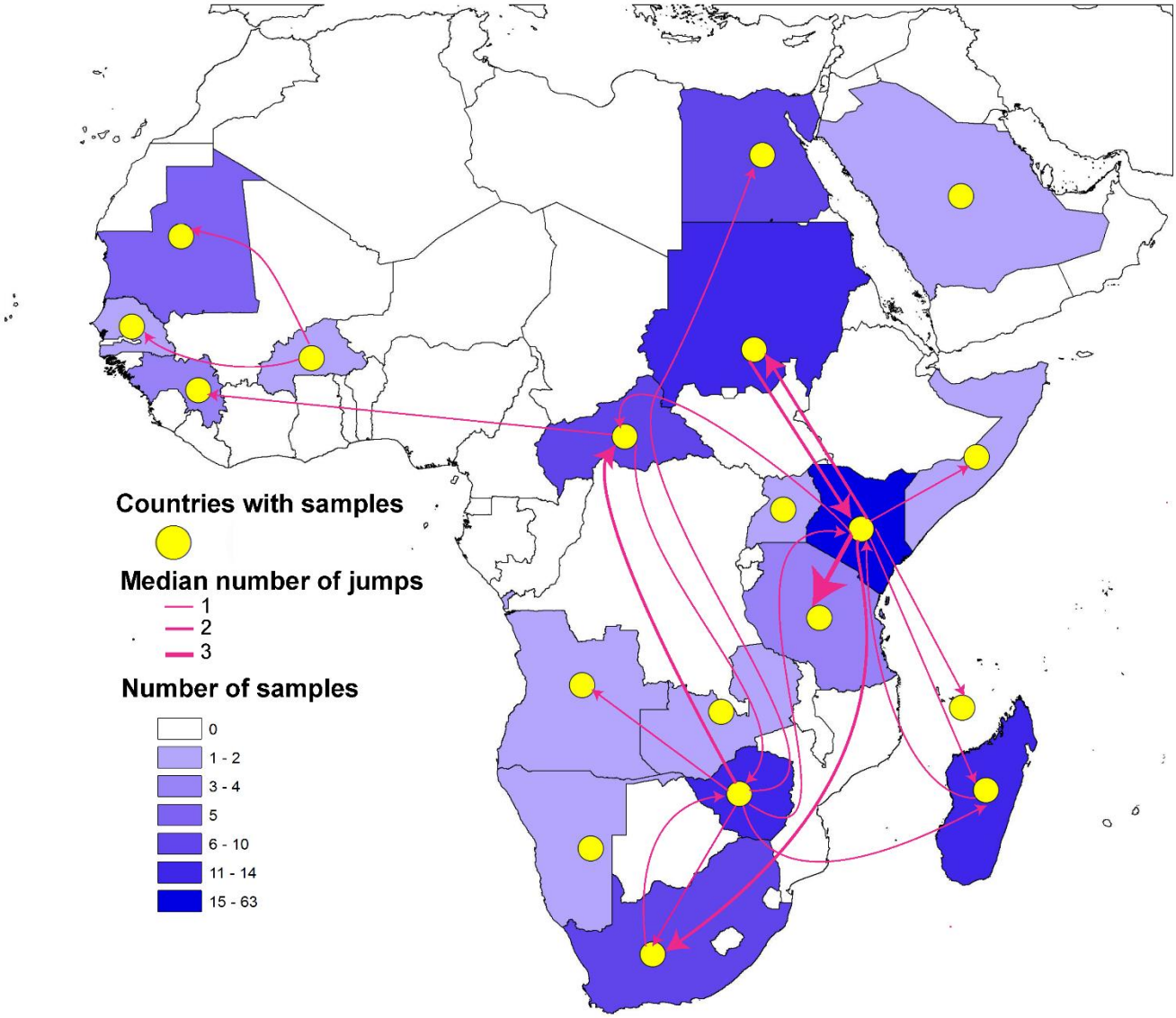
S3 File: Maximum Clade Credibility tree based on the small minigenomic segment (S) of RVFV. Country of origin is indicated by color on the tree branches and branch tips.



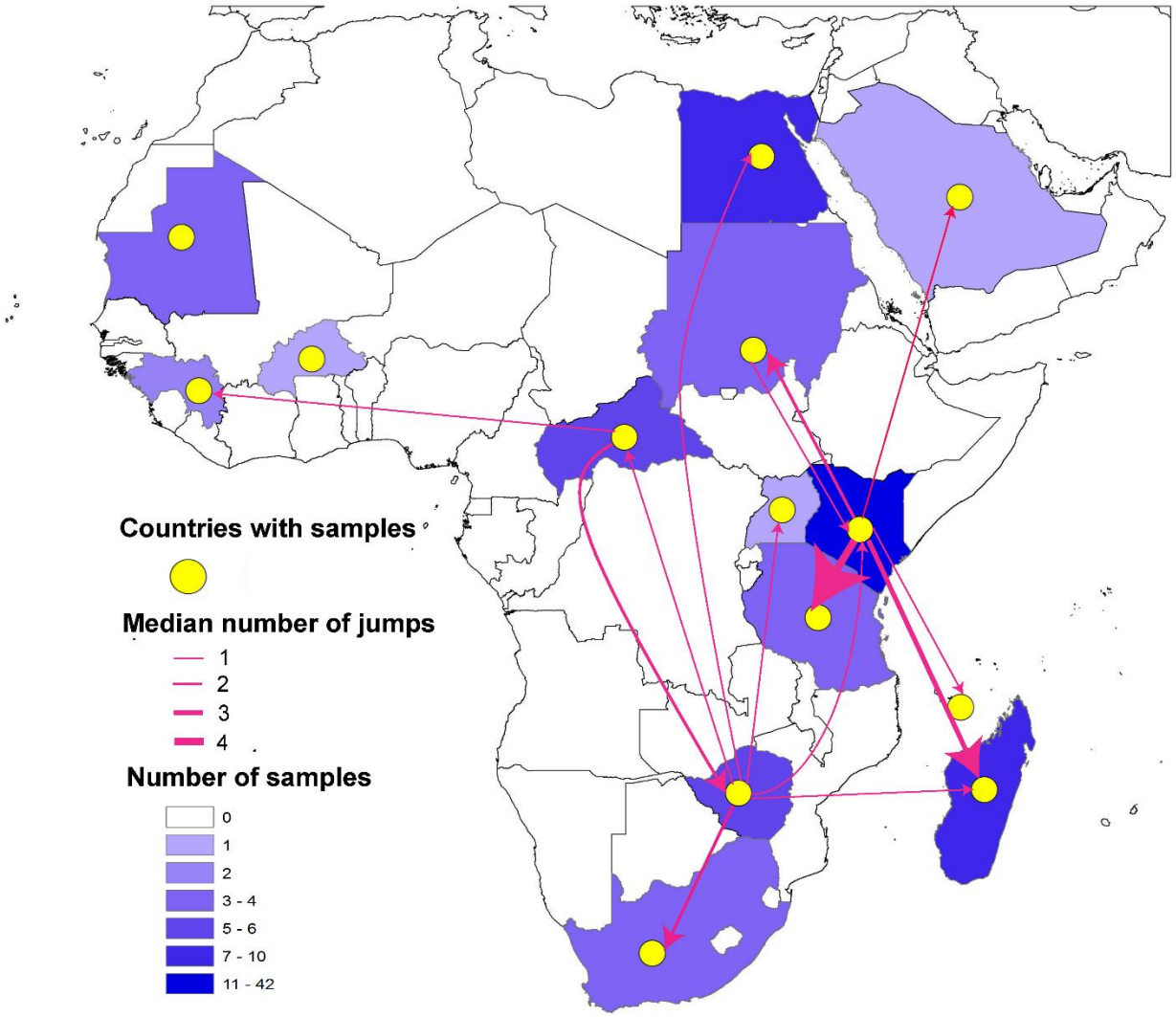
S4 File: Maximum Clade Credibility tree based on all sequences of the large minigenomic segment (L) of RVFV isolates in the study. Country of origin is indicated by color on tree branches and branch tips.



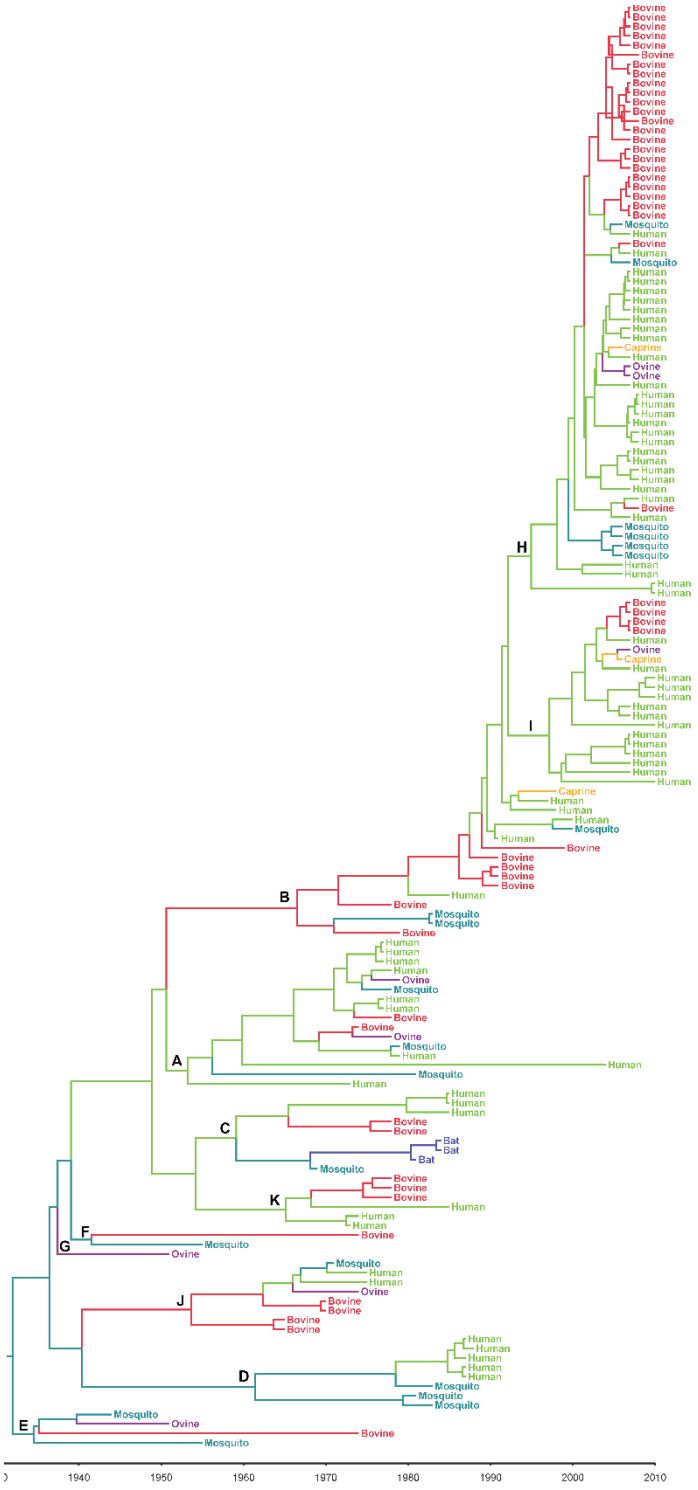
S5 File: Connectedness of countries with Rift Valley Fever outbreaks based on Markov Jumps analysis of the small minigenomic segment (S) of RVFV isolates in the study. The map shows only countries with non-zero transition frequencies. Connections between countries are presented as lines with arrows to refer to the direction of movement. Line thickness identifies the median number of jumps between each country pair.



S6 File: Connectedness of countries with Rift Valley Fever outbreaks based on Markov Jumps of the large minigenomic segment (L) of RVFV isolates in the study. The map shows only countries with non-zero transition frequencies. Connections between countries are presented as lines with arrows to refer to the direction of movement. Line thickness identifies the median number of jumps between each country pair.



S7 File: Maximum Clade Credibility tree based on the small minigenomic segment (S) of RVFV isolates in the study. The color of the tree branches and branch tips identifies the host of RVFV strains.



S8 File: Maximum Clade Credibility tree based on all sequences of the large minigenomic segment (L) of RVFV isolates in the study. The color of the branches and branch tips indicates the host of RVFV strains.